

PSYCHOLOGICAL REPORTS

ISSN 0033-2941

October 2013

PLD
Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221



PSYCHOLOGICAL REPORTS

Psychological Reports is published bimonthly, two volumes a year, the first with issues in February, April, and June, and the second with issues in August, October, and December, from P.O. Box 9229, Missoula, Montana 59807-9229. Subscriptions are accepted only for full calendar years. A minimum of 1,920 pages is planned for 2014. For subscription information, please visit www.AmSci.com and click on "Subscribers."

The purpose of this journal is to encourage scientific originality and creativity in the study of experimental, theoretical, and social psychology including: mental health; rehabilitation; intelligence and creativity; development and ageing; motivation; burnout; substance abuse; coping; communication; suicide and criminal behavior; organizational and consumer behavior; sports psychology; personality; fitness and health behavior. New or translated psychometric instruments may be described. Comments and special reviews are occasionally accepted. Controversial material of scientific merit is welcomed. Submitted manuscripts are all subject to rigorous peer review by outside experts chosen for their knowledge in the particular topic and/or general expertise in design, method, and analysis. In addition, associate editors who have broad knowledge of various topic areas are asked to review particularly difficult, unique, or controversial manuscripts. All manuscripts have at least 3 and up to 20 peer reviewers. Critical editing is combined with specific suggestions from multiple referees of each paper to help authors meet standards. For instructions for submitting a manuscript, please visit our web site (www.AmSci.com) and click on "Authors."

If a manuscript is accepted for publication, there are three possible formats for publication.

Regular articles. These are articles which require three or more printed pages.

One- to three-page "brief articles." This arrangement is useful if a particular finding can be reported completely in this format and will be immediately useful to other researchers in the field. Charge is \$35.00 per printed page plus compositor fees for tables, figures, or equations. Authors may submit supplemental material for filing with the Archive for Psychological Data.

Monograph supplements. Certain papers printing to more than 50 pages are published as monograph supplements. These are distributed to subscribers as parts of regular issues and are also made available as separates.

Publication is in order of receipt of proof from the authors. Author fees are \$35.00 per page, plus costs of special composition, e.g., tables, figures, and mathematics. Fees are paid by authors or their institutions.

This journal is supported by a fully searchable electronic database at www.AmSci.com. All articles and reference metadata are submitted to CrossRef for linking to other publishers and databases, so that the materials are accessible to online literature search. It is also the policy of this journal to file raw data with the Archive for Psychological Data. Authors should submit appropriate tables with their articles.

Responsibility for address change rests with the subscriber. Claims for missing issues must be made within two months of publication.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941

Vol. 113, No. 2

October 2013

SENIOR EDITORS

R. B. Ammons and C. H. Ammons

*Ammons
Scientific, Ltd.*

EDITORS

Douglas Ammons and S. A. Isbell

ASSISTANT EDITORS

Raymond Chester Russ and David Hecht

ASSOCIATE EDITORS

Charles I. Abramson
Oklahoma State University

Janet Kay Bobo
Battelle Seattle Research Center

Gerald L. Borofsky
Bay State Psychological Assoc., Inc.

James N. Butcher
University of Minnesota

Jean-Charles Chebat
HEC Montréal School of Management

Ross H. Day
La Trobe University, Bundoora

David N. Elkins
Pepperdine University

Norman H. Freed
Palo Alto, CA

Ronald E. Goldsmith
Florida State University

William M. Grove
University of Minnesota, Twin Cities

J. Ray Hays
Baylor College of Medicine

James C. Kaufman
California State University-San Bernardino

Richard C. Marken
Antioch University and UCLA

Gilbert W. Meier
National Institutes of Health

Gerald M. Meredith
University of Hawaii, Manoa

Peter F. Merenda
University of Rhode Island

Peter Mudrack
Kansas State University

Thomas M. Nelson
University of Alberta

Sherman Ross
Howard University

Walter R. Schumm
Kansas State University

E. Robert Sinnett
Kansas State University

Martin Voracek
University of Vienna

Cynthia M. Whissell
Laurentian University

Akiko Yamasaki
Kyoto University

Published bimonthly by Psychological Reports (USPS 448-780), P.O. Box 9229, Missoula, Montana; printed by Ammons Scientific, Ltd., 1917 South Higgins Avenue, Missoula, Montana 59801-6764. Please address correspondence and changes of address to P.O. Box 9229, Missoula, Montana 59807-9229. Periodicals Postage Paid at Missoula, Montana. POSTMASTER: Send address changes to *Psychological Reports*, P.O. Box 9229, Missoula, Montana 59807-9229.

Copyright 2013

Psychological Reports

PERMISSION TO COPY MATERIAL IN THIS JOURNAL

It is the policy of this journal to obtain and hold the copyrights for all articles published, with the exception of a very few articles or parts of articles in the public domain. Authors are allowed and encouraged to make copies of their articles for any scholarly purpose of their own. Authors may post, in electronic format, the last pre-submission version of their manuscript, with full citation and a link to the e-Journal on the title page. Other digital format posting of published articles is not permitted. Other scholars are hereby given permission to make single or multiple copies of single articles for personal scholarly use without explicit written permission, providing that no such copies are sold or used for commercial purposes. We consider it the obligation of any user to acknowledge sources explicitly and to obtain permission of the authors if extensive use is to be made of materials or information contained in an article. Extensive use of copy privileges in such a way as to substitute for a subscription to this journal is not intended or permitted. Such use would be unethical and, in the long run, self-defeating, since it would destroy the very basis for publication.

These policies are proposed on a trial basis and will be followed until further notice.

The Editors

July 15, 2004

Psychological Reports has been printed on acid-free paper since 1962, Vol. 10, No. 1 (February).

DISCLAIMER

The statements and opinions expressed in this Journal by its contributors, authors, and others are solely their own and are not necessarily those of the Editors. Publication of these ideas and expressions should not be viewed as endorsement. Contributing authors are expected to follow ethical procedures and sound scientific principles. Scientific inquiry is a process comprising discovery, interpretation, and organization of factual information. The Journal and its Editors cannot be held responsible for any detriment that may result from anyone's reliance on materials published in this Journal, since they represent only a fraction of the ongoing research in psychology and its topic areas. The published material which is supplied herewith is provided "As is." While we have taken reasonable measures to ensure accuracy, neither the publisher nor anyone else makes any warranties of any kind, either express or implied, including the limited warranties of merchantability, quality, accuracy, or fitness for any particular purpose.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Disability &
Trauma*



TARGET-TO-DISTRACTOR RATIO EFFECTS ON DECISION TIME IN THE ORDERLY ARRAY SHAPE CANCELLATION TASK^{1,2}

YOSHIMI NAKAJIMA, YOSHIFUMI IKEDA, AND HIDEYUKI OKUZUMI

Tokyo Gakuhei University

Summary.—The cancellation task is a paper-and-pencil test commonly used to assess attention or planning. This study investigated whether the decision time performance on the task was influenced by the number of targets and distractors. This study reduced the demand of planning and used an index of the decision time, an estimate of the time taken to decide whether to mark a stimulus. Forty healthy adults (M age = 21.3 yr., SD = 1.5) performed five cancellation tasks. Four tasks were conducted with instructions to mark a detected target. The target-to-distractor ratio varied from 35/15, 40/10, 45/5, and 50/0, and one task with instructions to mark all stimuli (50/0), to measure the motor time to mark targets. One-way analysis of variance indicated statistically significant differences between conditions. There was a linear relationship between decision time and target-to-distractor ratio; the decision time increased as the proportion of distractors increased. The results suggested the decision time reflects the frequency of switches between responses to targets and distractors or attention modulation of processing to targets and distractors.

Cancellation tasks are paper-and-pencil tests that require placement of a mark through each occurrence of a specific target displayed in an array of distractors on a sheet of paper, as quickly and accurately as possible. Successful performance on cancellation tasks requires several cognitive abilities, including sustained and selective attention, visual search, psychomotor speed, and fine motor coordination (Byrd, Touradji, Tang, & Manly, 2004). Performance is usually measured as number of errors, position of errors, visual search patterns, and time to complete the task.

Cancellation tasks have been used frequently for many years in neuropsychology as sensitive measures for detecting patients with visual neglect, whose performance is characterized as an absence of cancellation for either side of the test sheet (Weintraub, & Mesulum, 1988; Vanier, Gauthier, Lambert, Pepin, Robillard, Dubouloz, *et al.*, 1990; Manly, Dove, Blows, Melanie, Noonan, Teasdale, *et al.*, 2009). These tasks have also been used as an assessment of attentional ability in patients with traumatic brain injury, stroke, and Alzheimer's disease (Geldmacher & Hills, 1997; Amieva, Lafont, Dartigues, & Fabrigoule, 1999; Solfrizzi, Panza, Torres, Capurso, D'Introno, Colacicco, *et al.*, 2002; Nurmi, Kettunen, Laihosalo, Ruuskanen, Koivisto, & Jehkonen, 2010).

¹Address correspondence to Yoshimi Nakajima, Department of Special Needs Education, Faculty of Education, Tokyo Gakuhei University, Nukuikita-machi, Koganei-shi, Tokyo 184-8501, Japan or e-mail (nakayoshi783@gmail.com).

²The authors thank all participants, and also Yosuke Kita, who assisted with this study.

Some recent studies have used cancellation tasks to investigate individual differences in healthy adults without visual neglect. Some studies have observed the differences in performance by age, educational level, or race (Byrd, *et al.*, 2004; Brucki & Nitrini, 2008; Huang & Wang, 2008, 2009; Warren, Moore, & Vogtle, 2008). Others have examined the types of search strategies used for completing the tasks. Their results show that healthy adults executed visual searches in an organized and systematic way, e.g., in a structured linear pattern in a left-to-right and top-to-bottom direction (Mort & Kennard, 2003; Warren, *et al.*, 2008). Deviations from this approach might signal disorders of planning or organization (Lowery, Ragland, Gur, Gur, & Moberg, 2004).

Several factors influence performance on the cancellation task. Weintraub and Mesulum (1988) reported that performance is affected by stimulus material (letters vs. shapes) and characteristics of the stimulus array (random vs. orderly). Geldmacher (1998) found a visual pop-out effect of the targets, reporting higher performance, measured as accuracy and time, when a target was the letter "O" than when it was the letter "I", when the letter "L" served as a distractor. Other reports described the effect of target-to-distractor ratio (Kaplan, Verfaellie, Meadows, Caplan, Pessin, & DeWitt, 1991; Chatterjee, Mennemeier, & Heilman, 1992; Geldmacher, 1996; Huang & Wang, 2009). These studies reported lower accuracy and increased time as the proportion of distractors increased, although it was not clear that the relationship was linear because it compared only two types of target-to-distractor ratio.

Presumably, attentional (top-down) control is related to the effect of the target-to-distractor ratio. Within a sequence of trials in a conflict task, such as the 'Go/No-go' test in which participants are required to give a simple response to task-relevant information (Go stimuli) and to withhold any response to task-irrelevant information (No-go stimuli), task-relevant information is processed faster when task-irrelevant information is rare (Logan & Zbrodoff, 1979; Nieuwenhuis, Yeung, Wildenberg, & Ridderinkhof, 2003; Abrahamse, Duthoo, Notebaert, & Risko, 2013). Because the cancellation task comprises both task-relevant information (targets) and task-irrelevant information (distractors), the cancellation task can be seen as a conflict task. Therefore, a higher proportion of targets should facilitate identification and processing of targets, and the performance time should decrease (i.e., faster decisions to cancel). However, a higher proportion of distractors should facilitate identification and processing of distractors, which leads to inhibition and slower performance. In earlier studies, however, these patterns in performance were not found, likely because performance was often affected by several factors such as attention and planning, because participants were not instructed in how to search the

stimuli, even in tasks with an orderly array (Casco, Tressoldi, & Dellantonio, 1998; Lowery, *et al.*, 2004; Warren, *et al.*, 2008). Further investigation is necessary to elucidate how attentional control affects performance in the letter-cancellation task.

This study was conducted to test the above assumptions about the effect of the target-to-distractor ratio on cognitive processing, while controlling the factors that may confound the relation. An orderly array task was used to reduce the necessity of planning. Participants were instructed to search for targets in a specific manner. Several tasks with various target-to-distractor ratios were designed, in which the targets were always more frequent than distractors. To examine how the target-to-distractor ratio, but not the number of stimuli, affects the performance, the number of total stimuli was equalized across task conditions. To control the motor time required to mark targets across task conditions, a control task was administered, in which participants marked all the letters presented. The time needed to decide whether the letters were real targets or not (decision time) could then be calculated by subtracting the motor time from the time to complete the task. The decision time was expected to increase as the proportion of distractors increased, because frequent occurrence of targets with distractors affects attentional control by facilitating the processing of distractors and thereby weakening the processing of targets (Logan & Zbrodoff, 1979; Nieuwenhuis, *et al.*, 2003; Abrahamse, *et al.*, 2013).

Hypothesis 1. There would be a significant increase in decision time with decreasing target-to-distractor ratio.

Hypothesis 2. The increase in decision time should be linear as the number of distractors increases.

METHOD

Participants

The participants in this study were 40 university students (20 men, 20 women; $M = 21.3$ yr., $SD = 1.5$, range = 18–24). They volunteered for the study after recruitment from universities in Tokyo, Japan. Only right-handed participants were retained because the viewing angle for the stimuli varies with the dominant hand. Participants self-reported that they were free of any physical illness, visual or motor problem, or perceptual or cognitive disorder. Informed consent was obtained from all participants.

Measures

A paper-and-pencil version of the orderly array cancellation task was administered. This study had four cancellation tasks with various target-to-distractor ratios: 35 targets and 15 distractors (Form 35/15, targets

70%), 40 targets and 10 distractors (Form 40/10, targets 80%), 45 targets and 5 distractors (Form 45/5, targets 90%), and 50 targets and 0 distractors (Form 50/0, targets 100%). For these four cancellation tasks, participants were instructed to mark a target through each occurrence of a target. Before the test session, participants were not informed that there were no distractors on Form 50/0. In addition to these four tasks, participants were administered a motor task that used Form 50/0 with instructions to mark all stimuli.

Figure 1 shows the stimuli and test sheets for the cancellation task used for this study. Two geometric figures were used as the target and the distractor (Fig. 1A). For all five tasks, a total of 50 stimuli were arranged in a standard white A4 paper (21 × 29 cm). All stimuli were 2 cm high × 2 cm wide and placed in five lines, with a wide space of 1 cm and a vertical space of 2 cm open between stimuli (Fig. 1B).

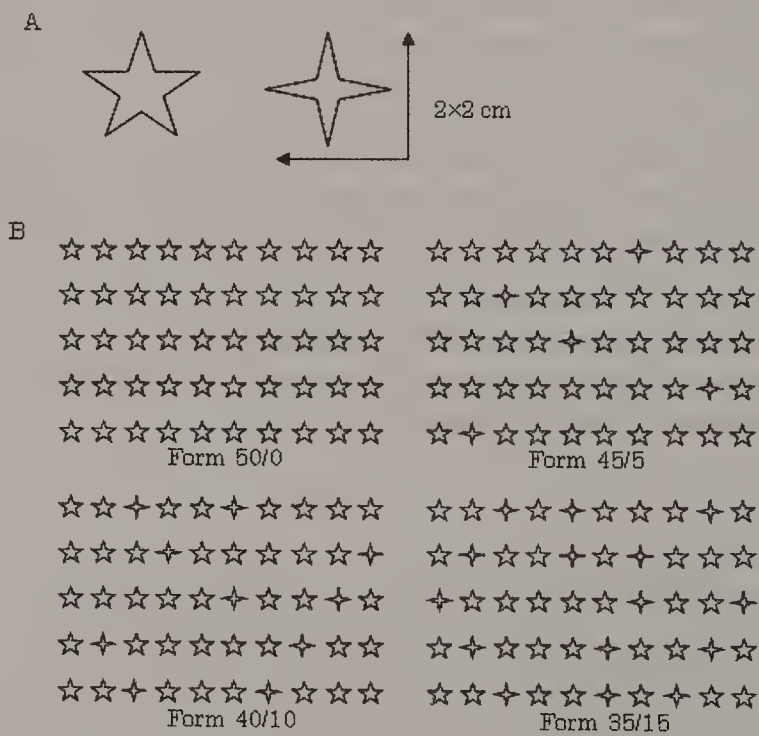


FIG. 1. Stimuli and test sheets in the cancelling task. A. Target shape (left) and distractor shape (right). B. Four cancellation forms.

Procedure

Participants were tested individually at a desk in a well-lit and quiet room at a university. The test sheets were placed on the desk, centered approximately 30 cm from the edge and at the participant's midline. Each

participant was instructed to use a pen with their dominant (right) hand to make a slash as quickly and accurately as possible through every occurrence of the target, but not to slash through any occurrence of the distractor. Each participant was instructed to use a specific searching path: a left-to-right and top-to-bottom pattern. Each participant was instructed not to go back to correct mistakes. The order of the four cancellation tasks (Form 35/15, Form 40/10, Form 45/5, and Form 50/0) was randomized. The motor task (Form 50/0 with instructions to mark every stimulus) was administered both before and after these four tasks.

Data Analysis

Items were counted as errors when a participant did not cancel a target (i.e., an omission error). No participant made a commission error. Numbers of errors and time to complete the task were recorded. The experimenter used a stopwatch to time the interval between a participant's first mark and the last mark for each task.

To evaluate the time taken to ascertain whether to mark the targets and whether to mark the distractors or not, the decision time was calculated using an original method. Firstly, the motor time to mark one target was calculated as the time to complete the motor task divided by the number of targets marked (50). Secondly, the motor time in each task condition was calculated as the product of the motor time to mark one target calculated above and the number of marks in each task. Finally, the decision time was calculated by subtracting the motor time for each task from the time to complete each task. This time can be described as "decision time = time to complete a task – (number of marks × time to complete a motor task / 50)."

Statistical Analysis

One-way analysis of variance (ANOVA) across the four target-to-distractor ratios was conducted to test Hypothesis 1, that there would be a statistically significant increase in decision time with increasing distractor ratio. To test Hypothesis 2, the coefficient of determination of the regression line of the decision time on the target-to-distractor ratio was examined. Curve fitting and statistical tests were performed with Excel.

RESULTS

Correct Response Rates and Time to Complete a Task

Table 1 shows means and standard deviations of correct response rates and time to complete a task for each cancellation task and motor task. In the cancellation tasks, the participants showed very accurate performance. Errors were few and not normally distributed. Therefore, analyses of errors were not conducted further. For the time to complete, one-way

TABLE 1
DESCRIPTIVE STATISTICS FOR CORRECT RESPONSE RATES AND TIME FOR FOUR FORMS
AND MOTOR TASK

Target-to-distractor Ratio	Correct Response Rate (max. = 1)		Decision Time, sec.		Total Time, sec.	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
35/15	0.998	0.010	3.44	1.33	14.86	3.84
40/10	1.000	0.000	3.00	1.41	16.06	4.51
45/5	0.999	0.005	2.22	1.51	16.90	5.04
50/0	0.999	0.006	1.92	1.20	18.25	5.35
Motor Task (50/0)	1.000	0.000			16.32	4.82

ANOVA showed a significant effect of the target-to-distractor ratio ($F_{3,117} = 59.29, p < .01, \eta^2 = 0.60$). *Post hoc* Bonferroni tests indicated there were significant differences between performance on Form 35/15 and Form 40/10, between Form 40/10 and Form 45/5, and between Form 45/5 and Form 50/0 ($ps < .05$).

Decision Time

Figure 2 presents means and standard deviations of decision times for each cancellation task. One-way ANOVA comparing the decision time across the four task levels indicated a significant effect of the target-to-distractor ratio ($F_{3,117} = 22.32, p < .01, \eta^2 = 0.36$). *Post hoc* Bonferroni tests indicated a significant difference between Form 40/10 and Form 45/5 ($p < .05$). No other comparison between forms was statistically significant. A linear functional relation was found between the decision time and the target-to-distractor ratio ($y = -0.53x + 3.98, R^2 = .97$).

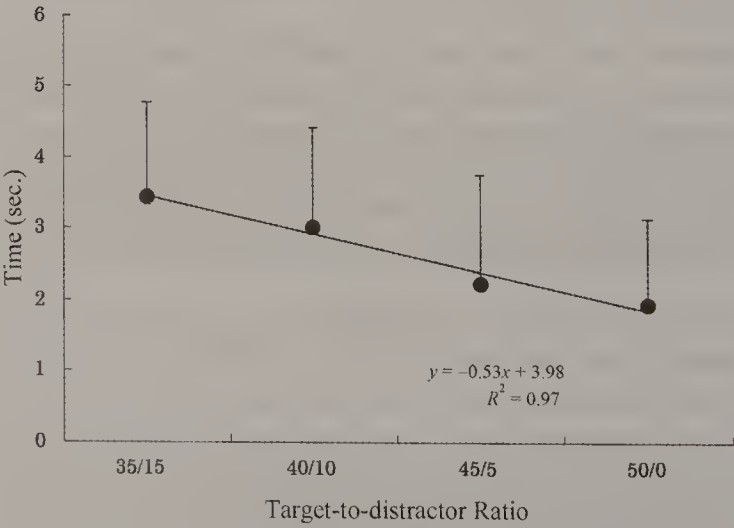


FIG. 2. Regression of decision time on target-to-distractor ratio.

DISCUSSION

This study was conducted to investigate whether the decision time, as an index of the performance in the cancellation task, was influenced by the target-to-distractor ratio. It was hypothesized that the decision time increases as the proportion of distractors increases because frequent occurrence with distractors affects attentional control, facilitating the detection of targets as their frequency becomes greater, and alternatively, weakening the processing of the targets as their frequency decreases (Logan & Zbrodoff, 1979; Nieuwenhuis, *et al.*, 2003; Abrahamse, *et al.*, 2013). Results indicated a strong relation between the decision time and the target-to-distractor ratio. The decision time tended to decrease as the proportion of targets increased, although a significant difference was found only between Form 40/10 and Form 45/5. In other words, the decision time tended to increase as the proportion of distractors increased. In addition, a linear functional relation was found between the decision time and the target-to-distractor ratio.

The relation between the decision time and the target-to-distractor ratio might reflect the frequency of switches between responses to targets and distractors. This study used target-to-distractor ratios in which targets always occurred more frequently than distractors. Participants might have anticipated the occurrence of targets where the targets were frequent. However, participants were required to switch task sets between a target and a distractor every time they encountered a distractor. A higher proportion of distractors might increase the demand for switching. If this is the case, then future investigations can be expected to reveal a nonlinear functional relation between the decision time and the target-to-distractor ratio for widely varied target-to-distractor ratios, including where targets were always more frequently occurring than distractors and where targets were always less frequently occurring than distractors. The inverted V-shaped relation with the decision time of Form 25/25 at the top would be expected.

However, the relation between the decision time and the target-to-distractor ratio might reflect modulation of processing to targets and distractors by attentional control. Responses to the targets can be regarded as the 'go' response because they require motor responses. In contrast, responses to the distractors can be regarded as a 'no-go' response because they require withholding of motor responses. Within a sequence of trials in a conflict task, such as the 'Go/No-go' task, non-conflicting stimuli are processed faster when conflicting stimuli are rare (Logan & Zbrodoff, 1979; Abrahamse, *et al.*, 2013). If this is the case, then future investigations can be expected to reveal a linear functional relation between the decision time and the target-to-distractor ratio in the target-to-distractor ratios where targets are always less frequent compared to distractors.

Results of this study revealed a linear functional relation between the decision time and the target-to-distractor ratio in the cancellation task where the targets were 70%, 80%, 90%, and 100% of all stimuli. The cancellation task used for this study might be useful to examine attentional control such as a switching cost and attention modulation (Abrahamse, *et al.*, 2013).

This study did not examine cancellation tasks in target-to-distractor ratios where targets occurred less frequently than distractors, e.g., Form 5/45, Form 10/40, and Form 15/35. Future studies including these target-to-distractor ratios are expected to reveal which mechanisms are involved strongly.

REFERENCES

- ABRAHAMSE, E. L., DUTHOO, W., NOTEBAERT, W., & RISKIO, E. F. (2013) Attention modulation by proportion congruency: the asymmetrical list shifting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. DOI: 10.1037/a0034060 (e-pub before print)
- AMIEVA, H., LAFONT, S., DARTIGUES, J. F., & FABRIGOULE, C. (1999) Selective attention in Alzheimer's disease: analysis of errors in Zazzo's Cancellation Task. *Brain and Cognition*, 40, 26-100.
- BRUCKI, S. M. D., & NITRINI, R. (2008) Cancellation tasks in very low educated people. *Archives of Clinical Neuropsychology*, 23, 139-147.
- BYRD, D. A., TOURADJI, P., TANG, M. X., & MANLY, J. J. (2004) Cancellation test performance in African American, Hispanic, and White elderly. *Journal of the International Neuropsychological Society*, 10, 401-411.
- CASCO, C., TRESSOLDI, P. E., & DELLANTONIO, A. (1998) Visual selective attention and reading efficiency are related in children. *Cortex*, 34, 531-546.
- CHATTERJEE, A., MENNEMEIER, M., & HEILMAN, K. M. (1992) A stimulus-response relationship in unilateral neglect: the power function. *Neuropsychologia*, 30, 1101-1108.
- GELDMACHER, D. S. (1996) Effect of stimulus number and target-to-distractor ratio on letter cancellation tasks. *Brain and Cognition*, 32, 405-415.
- GELDMACHER, D. S. (1998) Stimulus characteristics determine processing approach on random array letter-cancellation tasks. *Brain and Cognition*, 36, 346-354.
- GELDMACHER, D. S., & HILLS, E. C. (1997) Effect of stimulus number, target-to-distractor ratio, and motor speed on visual spatial search quality following traumatic brain injury. *Brain Injury*, 11, 59-66.
- HUANG, H. C., & WANG, T. Y. (2008) Visualized representation of visual search patterns for visuospatial attention test. *Behavior Research Methods*, 40, 383-390.
- HUANG, H. C., & WANG, T. Y. (2009) Stimulus effects on cancellation task performance in children with and without dyslexia. *Behavior Research Methods*, 41, 539-545.
- KAPLAN, R. F., VERFAELLIE, M., MEADOWS, M. E., CAPLAN, L. R., PESSIN, M. S., & DEWITT, L. D. (1991) Changing attentional demands in left hemispatial neglect. *Archives of Neurology*, 48, 1263-1266.
- LOGAN, G. D., & ZBRODOFF, N. J. (1979) When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, 7, 166-174.

- LOWERY, N., RAGLAND, J. D., GUR, R. C., GUR, R. E., & MOBERG, P. J. (2004) Normative data for the symbol cancellation test in young healthy adults. *Applied Neuropsychology*, 11, 216-219.
- MANLY, T., DOVE, A., BLOWS, S., MELANIE, G., NOONAN, M. A. P., TEASDALE, T. W., DODDS, C. M., FISH, J., & WARBURTON, E. (2009) Assessment of unilateral spatial neglect: scoring star cancellation performance from video recordings—method, reliability, benefits, and normative data. *Neuropsychology*, 23, 519-528.
- MORT, D. J., & KENNARD, C. (2003) Visual search and its disorders. *Current Opinion in Neurology*, 16, 51-57.
- NIEUWENHUIS, S., YEUNG, N., WILDENBERG, W., & RIDDERINKHOF, K. R. (2003) Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 17-26.
- NURMI, L., KETTUNEN, J., LAIHOSALO, M., RUUSKANEN, E.-I., KOIVISTO, A.-M., & JEKKONEN, M. (2010) Right hemisphere infarct patients and healthy controls: evaluation of starting points in cancellation tasks. *Journal of the International Neuropsychological Society*, 5, 902-909.
- SOLFRIZZI, V., PANZA, F., TORRES, F., CAPURSO, C., D'INTRONO, A., COLACICCO, A. M., & CAPURSO, A. (2002) Selective attention skills in differentiating between Alzheimer's disease and normal aging. *Journal of Geriatric Psychiatry and Neurology*, 15, 99-109.
- VANIER, M., GAUTHIER, L., LAMBERT, J., PEPIN, E. P., ROBILLARD, A., DUBOULOZ, C. J., GAGNON, R., & JOANNETTE, Y. (1990) Evaluation of left visuospatial neglect: norms and discrimination power of two tests. *Neuropsychology*, 4, 87-96.
- WARREN, M., MOORE, J. M., & VOGTLE, L. K. (2008) Search performance of healthy adults on cancellation tests. *American Journal of Occupational Therapy*, 62, 588-594.
- WEINTRAUB, S., & MESULUM, M. M. (1988) Visual hemispatial inattention: stimulus parameters and exploratory strategies. *Journal of Neurology, Neurosurgery, and Psychiatry*, 51, 1481-1488.

Accepted September 27, 2013.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Employment
Psychology &
Marketing*



VALIDATION OF THE KARASEK-JOB CONTENT QUESTIONNAIRE TO MEASURE JOB STRAIN IN VIETNAM¹

THI GIANG HOANG

*Faculty of Public Health, Hai Phong Medical University, Hai Phong,
Vietnam Institut de la Francophonie pour la Médecine tropicale, Vientiane, Lao PDR*

MARC CORBIÈRE

*Department of Rehabilitation, Université de Sherbrooke, Longueuil, Canada
Centre for Action in Work Disability Prevention and Rehabilitation (CAPRIT)*

ALESSIA NEGRINI

Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), Montréal, Canada

MINH KHUÊ PHAM

Faculty of Public Health, Hai Phong Medical University, Hai Phong, Vietnam

DANIEL REINHARZ

*Department of Social and Preventive Medicine, Laval University, Quebec, Canada
Institut de la Francophonie pour la Médecine tropicale, Vientiane, Lao PDR*

Summary.—The objective of this study was to validate the Karasek-Job Content Questionnaire in Vietnamese. A translation/back-translation of the questionnaire was performed prior to its administration to 344 health personnel in Vietnam. Several psychometric properties of the Vietnamese version of the Karasek-Job Content Questionnaire were assessed. A valid Vietnamese version of the Karasek-Job Content Questionnaire was produced, composed of five subscales based on the original theoretical model: Psychological demands, Social support at work, Decision latitude-Autonomy, Decision latitude-Authority, and Skill discretion. Internal consistency and reliability coefficients for each subscale of the questionnaire were satisfactory. The correlations with depression and work absence indicators were weak but statistically significant, as expected. The Vietnamese version of the Karasek-Job Content Questionnaire will help Vietnamese researchers and clinicians appropriately evaluate and document the job strain of workers in Vietnamese workplaces.

Job strain is defined by a series of reactions that occur when workers are faced with a disparity between demands at work and their knowledge, skills, and aptitudes (Leka, Griffiths, & Cox, 2004). In the workplace, job strain or stress at work is mainly caused by poor work organization,

¹Address correspondence to Marc Corbière, Université de Sherbrooke, Faculty of Medicine and Health Sciences, Department of Rehabilitation, Centre for Action in Work Disability Prevention and Rehabilitation (CAPRIT), 150 Place Charles-Lemoyne, Longueuil, Quebec, Canada J4K 0A8, or e-mail (marc.corbiere@usherbrooke.ca).

such as the way tasks and systems are designed and managed. Job strain can also result when workers do not receive sufficient support from their colleagues and supervisors, and it can be exacerbated by a lack of support from one's social environment (Elo, Leppänen, & Jahkola, 2003; Leka, *et al.*, 2004; Département Epidémiologie en Entreprise, 2006; Borteyrou, Truchot, & Rasclé, 2009).

Job strain can lead to various mental and physical illnesses such as depression, metabolic disorders, hypertension, angina pectoris, and musculoskeletal disorders. These illnesses also affect companies, due to reduced performance by stressed workers, absenteeism and increased staff turnover, along with potential increased costs due to a greater risk of accidents at work (Leka, *et al.*, 2004; World Health Organisation, 2008).

In developing countries, little information exists about the prevalence of job strain. Nevertheless, it is estimated that 50% to 70% of workers (compared with 30% in industrialized countries) are forced to suffer or work harder due to poor ergonomic working conditions. Most at risk are miners, farmers, fishermen, loggers, workers in the construction sector, and health workers (World Health Organisation, 2008). Consequently, job strain is a public health problem, particularly in developing countries.

Vietnam is a middle-income country where some efforts have been made to address the issue of occupational illnesses but, to date, little attention has generally been paid to workers' mental health, particularly to job strain. There are several reasons for this neglect, such as a lack of trained human resources in occupational mental health or, more fundamentally, the government's lack of awareness about mental health-related occupational interventions. One factor contributing to this lack of awareness is the absence of a measure for stress at work or job strain, validated in the Vietnamese language, which would allow non-psychiatric health workers to quickly diagnose stress or job strain among employees. Such measures would provide better assessment of the problem and would support the development, implementation, and monitoring of prevention and treatment interventions for stress-induced health problems.

One of the main theoretical models used to explain the relationship between job strain and worker health is the job demand-control model (JD-C Model; Karasek, 1979), extended to the job demand-control-support model (Johnson & Hall, 1988). In the job demand-control model, workers who simultaneously experience high psychosocial job demands and low control are more likely to develop stress-related health problems (e.g., Karasek, 1979; Karasek & Theorell, 1990; Stansfeld & Candy, 2006; Nakao, 2010). Job demands consist of mental requirements, time pressures, excessive workload, task interruption, intense concentration, and conflicting demands. Control (i.e., decision latitude) is the combination

of decision authority and autonomy to manage work time (e.g., taking part in decisions, having freedom in the way the work is done, setting the times for starting and finishing work), and skill discretion (e.g., opportunities to learn new things, to develop and use new skills, creativity, non-repetitive work activities).

In 1988, Johnson and Hall introduced the third component of the model: social support at work, which concerns the socio-emotional integration and trust among supervisor and co-workers, and the level of help and assistance provided by others when one is performing tasks (Karasek & Theorell, 1990). The basic assumption of the job demand-control model (Karasek, 1979) is that psychosocial risk factors (i.e., job demands) lead to job strain when job resources are lacking (i.e., low control, low social support).

From a practical point of view, according to Karasek's model, jobs can be classified into four groups (high-strain jobs and low-strain jobs; passive jobs and active jobs) located on two axes (strain axis and learning axis). For example, high-strain jobs have the worst combination of factors: high psychological demands, low control, and lack of social support. Working under these adverse conditions can directly increase the risk of developing physical (e.g., coronary health diseases, ulcers) and mental health problems (e.g., anxiety, depression, fatigue) and, indirectly, increase the rate of absenteeism from work due to job strain/illness, particularly for occupations with high percentages of women (e.g. Stansfeld & Candy, 2006; Bonde, 2008; Nieuwenhuijsen, Bruinvels, & Frings-Dresen, 2010).

Karasek (1985) operationalized his theory by suggesting the utilization of the Karasek-Job Content Questionnaire (K-JCQ). This questionnaire has become the gold standard for assessing job strain and stress in the workplace. It has been translated into and used in more than 22 languages and countries (e.g., Europe, America and Asia) (Li, Yang, Liu, Xu, & Cho, 2004; Hadi, Naing, Daud, & Nordin, 2006; Eum, Li, Jhun, Park, Tak, Karasek, *et al.*, 2007; de Araújo & Karasek, 2008; Maizura, Retneswari, Moe, Hoe, & Bulgiba, 2010). The K-JCQ is composed of different dimensions corresponding to potential job stressors such as psychological demands, decision latitude, and social support at work. The K-JCQ is widely used to assess the relationships between work-related psychosocial risk factors (e.g., job content, social relationship at work) and health outcomes (i.e., mental and/or physical diseases), as well as work-related psychosocial risk factors and work-related behaviours (e.g., absenteeism; Houtman, Kornitzer, De Smet, Koyuncu, De Backer, Pelfrene, *et al.*, 1999).

The main objective of this study was to validate the K-JCQ with Vietnamese health personnel working in a hospital. Specifically, this study encompassed the following basic components of a validation study: (1) construct validity using exploratory factor analyses to identify the con-

ceptual dimensions (or subscales) from the K-JCQ administered to health personnel in Vietnam; (2) internal consistency for each emerging subscale using Cronbach's alpha coefficients; (3) test-retest reliability using a Pearson correlation with a one-week period for the retest of the Vietnamese K-JCQ; and, finally, (4) criterion validity by assessing the correlations between all subscales of the Vietnamese K-JCQ and depressive symptoms as well as variables related to work absences.

METHOD

Procedure and Participants

One of the researchers (G. H.) introduced the study to potential participants, i.e., health personnel from all departments working in the second largest provincial general hospital in Hai Phong, in the north of Vietnam. Health personnel who agreed to participate in the study were offered an appointment for a 30- to 45-minute session to complete the questionnaires (sociodemographics and K-JCQ) and to participate in a face-to-face interview (based on DSM-IV) for depression assessment. The data collection occurred at the end of the participants' working day in a conference room, under conditions of confidentiality. Both the Hai Phong Medical University and Kienan Hospital ethics committees approved the study.

The study sample consisted of 344 health personnel from diverse departments in the hospital where the most represented were medicine (13.7%), surgery (11.9%), gynecology-obstetrics (11.1%), reanimation (10.2%), pharmacology (6.7%), and anesthesiology (6.4%). The participants included 183 nurses (54%) and 74 physicians (21%), and the rest of the sample derived from other health professions (e.g., pharmacy, radiology, dentistry). Female participants numbered 256 (74.4%), and the average age of all participants was 34.7 yr. ($SD = 10.5$); 280 were married (81.4%), and the average time worked in the hospital was 11.3 years ($SD = 10.6$).

Measures and Translation of the Karasek-Job Content Questionnaire

Two questionnaires were administered to the participants: (1) a sociodemographic questionnaire including questions related to work absences; and (2) the K-JCQ translated into Vietnamese. Also, a face-to-face interview was conducted to establish a diagnosis of depression and to obtain a depression score for participants (using the depression criteria of the DSM-IV).

Demographics.—The socio-demographic questionnaire included questions on sex, age, marital status, occupation, and time spent in the present job. Participants were asked additional questions about work absences over the previous three months: Have you been absent from work in the last three months (0: No, 1: Yes)? If yes, how many days were you absent

in the last three months (number of days)? If yes, could you please explain the causes of these absences? Was the reason for your absence due to working conditions (0: No, 1: Yes)?

Job content.—In this study, the 35-item format of the K-JCQ was used (Kompier & Levi, 1995). For each item, the response was recorded using a four-point Likert-type scale with anchors 1: Strongly disagree and 4: Strongly agree. As mentioned above, the K-JCQ is based on a conceptual model consisting of three conceptual dimensions: (1) psychological demands, i.e., constraints associated with the execution of a task (e.g., quantity, time constraints; nine items); (2) decision latitude, which includes the control one has over his work, i.e., the autonomy conceded by the employing organization, participation in decisions, and the level of skills to employ (18 items); (3) social support at work, i.e., the relationships with colleagues and supervisors (eight items). According to the Swedish Demand Control Support Questionnaire (DCSQ) proposed by Theorell, Perski, Akerstedt, Sigala, Ahlberg-Hulten, Svensson, *et al.* (1988), social support at work was considered to include the social and interpersonal characteristics of the work environment as a whole, without distinguishing between coworker or supervisor support, i.e., both stakeholders are considered in the same item, not separately.

As recommended for the transcultural validation of questionnaires (Vallerand, 1989), the K-JCQ was translated into Vietnamese by one bilingual researcher (G. H.) and then back-translated to French by a professional translator who was not involved in the study. Two of the team's researchers compared the original version of K-JCQ to the French back-translated version (French is one of the common foreign languages in Vietnam due to its colonial past). After a few modifications and adaptations to the Vietnamese language such as the term "work procedure" (Item 26), which required more precision (e.g., appointments with patients, to follow the tasks to do in the job), a second version was created (Table 1). A group of seven people working in the targeted hospital in different types of employment was then asked to read this second Vietnamese version of the K-JCQ and to comment on items that were difficult to read (formulation) and to understand (clarification). After minor modifications, this step led to the production of a final version of the K-JCQ in Vietnamese for use in the next phase, the validation of the questionnaire, i.e., construct validity, internal consistency, test-retest reliability, and criterion validity.

Depression.—The DSM-IV depression component is a diagnostic tool developed by the American Psychiatry Association that is widely used in Vietnam (Nguyen, 2003; Nguyen, 2013). This depression component contains nine criteria for diagnosing depression (e.g., problems with sleep, appetite), which are evaluated (presence or absence) over a two-

TABLE 1
ITEMS OF JOB CONTENT QUESTIONNAIRE STANDARDIZED INTO VIETNAMESE (VIETNAMESE VERSION)

Item	Translation
1	Công việc đòi hỏi tôi phải làm việc rất nhanh
2	Công việc buộc tôi phải làm việc rất vất vả
3	Họ bắt tôi làm việc quá mức
4	Tôi không đủ thời gian để hoàn thành công việc của mình
5	Tôi thường gặp phải những yêu cầu vô lý từ những người khác trong khi làm việc
6	Công việc đòi hỏi tôi phải tập trung cao độ trong thời gian dài
7	Công việc của tôi thường xuyên bị gián đoạn trước khi hoàn thành khiến sau đó tôi lại phải bắt đầu lại
8	Tôi luôn bận rộn trong công việc
9	Việc phải chờ đợi công việc từ những người khác hoặc bộ phận khác thường xuyên làm chậm công việc của tôi lại
10	Công việc cho phép tôi được tự quyết định nhiều thứ
11	Tôi được thoải mái quyết định xem mình sẽ làm việc như nào
12	Tôi có nhiều thứ để nói về những gì đang diễn ra trong công việc của mình
13	Tôi có thể ấn định được trình tự công việc mà tôi muốn thực hiện
14	Tôi có thể quyết định được thời điểm tiến hành công việc của mình
15	Tôi có thể dễ dàng rời bỏ công việc một thời gian ngắn
16	Tôi có thể tạm ngừng làm việc nếu tôi muốn
17	Tôi có thể tự quyết định nhịp độ công việc của mình
18	Công việc đòi hỏi tôi phải cập nhật những kiến thức mới
19	Công việc của tôi có rất ít các hoạt động lặp đi lặp lại
20	Công việc đòi hỏi tôi phải sáng tạo
21	Công việc của tôi đòi hỏi trình độ chuyên môn cao
22	Công việc của tôi bao gồm rất nhiều hoạt động
23	Tôi có cơ hội phát triển các khả năng của mình
24	Tôi có thể chọn được giờ bắt đầu và kết thúc công việc của mình
25	Tôi tự quyết định thời điểm nghỉ giải lao
26	Tôi tự quyết định ngày nghỉ phép
27	Tôi biết lộ trình công việc của mình trước ít nhất 1 tháng*
28	Bầu không khí nơi tôi làm việc rất tốt
29	Hiếm khi tôi và đồng nghiệp gây gổ
30	Nếu muốn, tôi có thể đề nghị được sự giúp đỡ từ đồng nghiệp
31	Tôi có mối quan hệ tốt với cấp trên của mình
32	Cấp trên của tôi rất lắng nghe ý kiến của tôi
33	Cấp trên của tôi luôn có một cái nhìn rõ ràng về cách thức mà tôi làm việc
34	Họ giúp đỡ tôi nhiều trong công việc
35	Tôi được thông tin đầy đủ về những gì đang diễn ra ở nơi làm việc

*This item corresponds to "I know my work procedure at least one month in advance (Item #27)," and it should be considered with special attention for future studies in Vietnam.

week period. According to the tool's guidelines, participants should be considered at risk for a depression diagnosis if they answered positively for five criteria out of nine depressive symptoms. In this study, the addition of participants' answers for each clinical symptom was calculated (0: Absence and 1: Presence) and expressed as an equivalent score of depressive symptoms (minimum 0, maximum 9). This depression score was used for assessing the criterion validity of the Vietnamese K-JCQ.

Analyses

Descriptive analyses (e.g., percentages, means) were calculated to describe the sample of the study as well as variables related to work absences. Construct validity was assessed using exploratory factor analysis. The principal components analysis enabled us to undertake factor extraction from the 35-item K-JCQ. Different indices were considered such as scree plot, eigenvalues, percentage of total variance, and item loadings on each emerging subscale. The sample size of 344 individuals was satisfactory for conducting factor analyses since the K-JCQ consists of 35 items (10 individuals for each item; Thompson, 1996). Cronbach's alpha was used to assess the internal consistency of each subscale found in the factor analysis. Then, to assess the test-retest reliability, the first 31 participants were asked to fill out the Vietnamese K-JCQ a second time after 5 to 7 days. A Pearson correlation was performed with scores obtained at Time 1 and Time 2 for each subscale emerging from the Vietnamese K-JCQ. Finally, criterion validity was estimated with statistically appropriate tests (Spearman Rho, Mann-Whitney *U* tests or Pearson correlations) to study the relationship between scores on the subscales of K-JCQ and the variables related to work absences.

RESULTS

Approximately one-third of the study population (31.7%) reported being absent from work at least once in the last 3 months (Table 2). This equaled 109 workers absent for an average of almost 5 days (cumulative absences; $SD = 14.6$ days). The main cause of absence given by workers was family problems (77.1%). When workers were asked if the absence was related to working conditions, 12.8% answered positively. Results from face-to-face interviews using the DSM-IV criteria showed that 4.7% ($n = 16$) of the sample were at high risk for having a depression diagnosis since they were positive for five criteria (clinical symptoms) out of nine for establishing the psychiatric diagnosis. The mean score ($n = 344$) for all nine depressive symptoms was 1.3 ($SD = 1.5$).

Results from the exploratory factor analysis indicated all eigenvalues were greater than 1.0 for five factors, supported by a break after the fifth factor (Scree plot). After Varimax rotation, the five factors had the fol-

TABLE 2
DESCRIPTION OF ABSENTEEISM AT WORKPLACE IN THE LAST 3 MONTHS (*N* = 109)

		<i>n</i> (%)	<i>M</i>	Median	Range	<i>SD</i>
Number of absences			1.39	1	1–4	0.74
Total number of days of absence			4.98	2	1–90	14.64
Causes	Family problems	84 (77.1)				
	Short-term disease	18 (16.5)				
	Maternity	3 (2.8)				
	Long-term disease	2 (1.8)				
	Traffic accident	2 (1.8)				
Absences due to perceived working conditions						
	yes	14 (12.8)				

lowing variance values, respectively: 11.3, 9.4, 8.8, 8.6, and 6.9 (Table 2), with 45% of the total of variance. From this final factor solution, it should be noted that one item, “I know my work procedure at least one month in advance,” was excluded because it did not load significantly ($< .30$) on one specific factor. All other items loaded significantly on their respective factors, ranging from .38 to .77 (Table 2). Table 2 shows that this factor solution was empirically and conceptually acceptable (five subscales). Interestingly, two subscales of the Vietnamese K-JCQ were identical to the original version of the tool: “Social support at work” and “Psychological demands.” Results from the Vietnamese sample indicated a division of the third dimension of the K-JCQ, entitled “Decision latitude,” into three subscales. The first two subscales referred to the Decision latitude dimension and were named: “Decision latitude–Autonomy” and “Decision latitude–Authority.” The third subscale consisted of items measuring the skills expected at work (e.g., variety, creativity). We referenced this to the original dimension and named it “Skill discretion.” At the test-retest of the questionnaires administered twice with an interval of a maximum of seven days, positive and significant correlation coefficients were obtained for all five factors (between .56 and .90), suggesting acceptable test-retest reliability of all subscales of the Vietnamese K-JCQ.

Table 3 presents the results of internal consistency analyses as well as the correlations with the score of depressive symptoms and variables of work absences. The five-factor solution from the exploratory factor analysis was supported by the acceptable internal consistency coefficients, ranging from .67 to .82. All subscales of the Vietnamese K-JCQ were either

TABLE 3
FACTOR SOLUTION OF JCQ-K ($N = 344$)

	1	2	3	4	5	h^2
Social support at work						
37. Immediate supervisor gives me enough support in my work	.77	-.11	.05	.19	-.01	.65
35. Immediate supervisor takes my ideas into account sufficiently	.76	.04	.06	.06	.07	.59
36. Immediate supervisor has a clear picture of how I work	.76	.02	-.06	-.06	.10	.59
38. I am sufficiently informed of what's happening at work	.73	.09	-.09	-.09	.01	.55
34. I have a good relationship with my immediate supervisor	.73	-.04	.13	.13	.14	.58
31. The atmosphere in the workplace is good	.58	.21	-.09	-.09	.06	.39
32. Aggressiveness is rare among my colleagues and me	.51	-.03	.03	.03	-.05	.27
33. If I want, I can get help from one or more colleagues	.42	-.17	.05	.05	-.07	.32
Decision latitude – Autonomy						
14. I can determine when to work	-.01	.62	-.06	-.05	.28	.46
25. I can interrupt my work as I wish	-.03	.62	.04	-.09	-.07	.39
16. I set break times myself	-.01	.62	-.08	-.09	.14	.41
15. I can easily leave work for short periods	-.04	.59	.15	-.10	-.30	.48
27. I determine days off myself	.08	.53	-.23	-.07	.09	.35
17. I can determine my own work pace	-.08	.52	-.08	.13	.33	.40
24. I can set the time when I start and finish my work	-.04	.48	-.09	-.15	.31	.36
19. My work includes some repetitive tasks	.04	.42	.09	.15	.01	.21
Psychological demands						
3. I'm required to do excessive work	-.07	-.03	.71	.05	.02	.51
2. My job requires me to work hard	.03	-.16	.68	.14	-.02	.50
6. My job requires long periods of intense concentration	-.07	-.19	.56	.18	.18	.42
8. I'm always in a hurry in my work	.17	-.11	.53	.18	.30	.45
4. I don't have enough time to finish my work	-.03	.17	.51	-.01	-.12	.31
7. My tasks are often interrupted before completion, which requires me to resume them later	-.11	.42	.51	.23	.07	.50
5. I'm exposed to conflicting demands from others	.07	-.11	.44	-.11	.28	.30
9. Requiring the work of other individuals or other services often slows me	-.01	.09	.42	.09	-.02	.19
1. My job requires me to work very quickly	.29	-.08	.38	.06	.05	.24

(continued on next page)

Note.—Boldface font indicates major saturations > .30. The last column corresponds to item communalities.

TABLE 3 (CONT'D)
FACTOR SOLUTION OF JCQ-K (*N* = 344)

	1	2	3	4	5	<i>h</i> ²
Skill discretion						
20. My job requires me to be creative	-.02	-.01	.06	.77	.05	.60
18. My job requires me to assimilate new knowledge	.15	-.16	.23	.69	.07	.59
21. My work involves a high level of qualification	.02	.05	.28	.67	.02	.52
23. I have the opportunity to develop skills	.08	.09	-.04	.66	-.02	.45
22. My work includes many activities	.11	-.08	-.02	.66	.15	.47
Decision latitude – Authority						
11. I have a lot of freedom to decide how I will do my job	.11	.23	-.03	-.06	.75	.63
12. I have much to say about what happens in my work	.05	-.08	.20	.15	.67	.52
10. My job allows me to make many decisions	-.02	.23	.06	.07	.63	.46
13. I can determine the order in which I perform my tasks	.15	.37	.06	.16	.43	.38
Eigenvalue	4.6	3.6	3.1	2.1	1.6	
Percent variance explained after rotation	11.3	9.4	8.8	8.6	6.9	

Note.—Boldface font indicates major saturations > .30. The last column corresponds to item communalities.

not correlated significantly or had low coefficients ranging from $r = .15$ to $r = .35$ ($p < .05$), indicating that this questionnaire measures several different aspects of job strain. In general, the highest (yet modest) correlation coefficients are those between the three factors (subscales) originally conceptualized to be within the same dimension, i.e., Decision latitude.

Finally, as shown in Table 4, there was a positive and statistically significant association between psychological demands and the depression score ($r = .18$, $p < .01$) as well as with the number of work absences ($r = .20$, $p < .05$). Interestingly, only the depression score correlated with the cause of work absences inherent to work conditions. To summarize, the greater the psychological demands on health personnel, the more often they present depressive symptoms, are absent from the workplace, and mention working conditions as the reason for their absence. We note that there was no statistically significant association between the total number of workday absences and subscale scores of the Vietnamese K-JCQ or between the total number of workday absences and the depression score.

DISCUSSION

The main goal of this study was to validate the K-JCQ (Karasek, 1979) in Vietnamese by considering different types of psychometric properties such as construct validity, internal consistency, test-retest reliability, and

TABLE 4
MEAN, STANDARD DEVIATIONS, ALPHAS AND CORRELATIONS BETWEEN SUBSCALES OF K-JCQ,
DEPRESSION SCORE, AND WORK ABSENCES VARIABLES

	1	2	3	4	5	6	7	8	9	10
K-JCQ										
1. Social support at work	.82(<i>M</i> = 3.0, <i>SD</i> = .3)									
2. Psychological demands	.08	.70(<i>M</i> = 2.4, <i>SD</i> = .3)								
3. Decision latitude—Autonomy	.00	-.05	.71(<i>M</i> = 2.3, <i>SD</i> = .4)							
4. Decision latitude—Authority	.15†	.18†	.35*	.67(<i>M</i> = 2.6, <i>SD</i> = .4)						
5. Skill discretion	.22†	.24†	-.07	.19*	.76(<i>M</i> = 2.9, <i>SD</i> = 0.4)					
6. Work absences (1 = Yes)	-.05	.00	-.12*	-.17†	-.07					
7. No. work absences	-.05	.20*	-.12*	.03	-.23*	(<i>M</i> = 1.4, <i>SD</i> = .7)				
8. Total no. workday absences	.17	-.02	.05	-.01	.08	-.05	(<i>M</i> = 5.0, <i>SD</i> = 14.6)			
9. Cause of absence related to work conditions (1 = Yes)	-.14	.08	.04	.01	.07	-.06	-.01			
10. Depression score	-.10	.18†	-.06	.04	.08	-.09	.11	.33†	(<i>M</i> = 1.3, <i>SD</i> = 1.5)	

Note.—Italic numbers on the diagonal are Cronbach's alpha coefficients. The scores for each K-JCQ subscale and score of depression were calculated by summing the score of each item, which was divided by the number of items. The sample size is *N* = 344 for all variables except for variables named: number of work absences, total number of workdays absence, and cause of absence related to work conditions, for which the sample size was *N* = 109. **p* < .05, †*p* < .01.

criterion validity. The rationale for choosing the K-JCQ for evaluating job strain in the Vietnamese workplace consists of at least two essential reasons. Firstly, to our knowledge no validated measures of job strain exist in Vietnam to assess intervention in this public health problem and, consequently, there is an empirical void to fill. Secondly, the K-JCQ is a well-established questionnaire worldwide, with more than 20 cultural and linguistic adaptations,² and the validation of this questionnaire in the Vietnamese context would facilitate international comparisons.

In assessing the construct validity, the exploratory factor analysis allowed the identification of a specific model for Vietnam, with five subscales including Psychological demands and Social support at work, both present in the original version of the K-JCQ (Kompier & Levi, 1995). With respect to the original version of the K-JCQ, the dimension entitled Decision latitude was expressed in the Vietnamese sample through three subscales called: Decision latitude–Autonomy, Decision latitude–Authority, and Skill discretion. In fact, different factor structures have been found and different subscales of measurement for the potential psychosocial risk factors at work have been validated (i.e., Pelfrene, Vlerick, Mak, De Smet, Kornitzer, & De Backer, 2001). For instance, Karasek, Brisson, Kawakami, Houtman, Bongers, and Amick (1998) measured employees' control on the job using a set of subscales: skill discretion (i.e., variety of skills and creativity required on the job); decision authority (i.e., the possibility for employees to make decisions about their work); skill underutilization (i.e., loss of acquired skills); and the macro-level component of decision latitude (i.e., high-level influence on group, union, and company). However, the macro-level dimension of decision latitude is rarely used, and it could be dropped from the questionnaire because the remaining items can properly measure the individual decision latitude, with the three and specific subscales emerging from the Vietnamese sample.

Furthermore, other validation studies have reported that the main factor solution of the dimension Decision latitude was divided into two subscales: Decision authority and Skill discretion. In general, these two subscales are described as empirically related but theoretically separated (Schreurs & Taris, 1998).² However, no consensus exists for the operationalization of job control (e.g., Schreurs & Taris, 1998; de Jonge, Reuvers, Houtman, Bongers, & Kompier, 2000; Pelfrene, *et al.*, 2001). This is because a theoretical difference exists between these two concepts: decision authority represents the specific individual behavioral control over work (i.e., autonomy and participation in the decision-making process); whereas, skill discretion consists of qualities related to the characteristics of the work (e.g.,

²See the website: <http://www.jcqcenter.org/>

complexity, creativity; Ganster & Perrewé, 2011). Taking this into account, the actual results confirm that the decision authority dimension is separated from skill discretion. At the same time, in the Vietnamese population, Decision latitude is measured by two subscales named: Decision latitude–Autonomy, referring to the possibility of managing and controlling working time, and Decision latitude–Authority, referring to the possibility of taking part in the decision making process and, more specifically, having control over the process of the work. Even though these two last subscales can have similarities, the former is related to the control of time and chronology, whereas the latter is related to making decisions, *per se*, on the job. In total, there are three distinct concepts (or subscales) regarding the larger conceptual dimension of latitude decision in Vietnamese culture. Consequently, these results indicated a clear distinction between the types of decision latitude in the Vietnamese context that could be partly explained by its sociopolitical history. Notwithstanding that the Vietnamese context for the decision latitude concept is very detailed with subtle nuances, it likely remains an emotionally-charged concept, considering its cultural and political history.

Also, linked to the exploratory factor analysis results, out of 35 items from the original version of the K-JCQ, only the item “I know my work procedure at least one month in advance” did not significantly load on any subscale of the five factor solution of the K-JCQ, leading to its removal from the Vietnamese version. The exclusion of this item comes either from the difficulty in accurately translating this item into the Vietnamese language and culture or the inadequacy of the content of this item for health personnel working in a professional context where it is difficult to establish work procedures a long time in advance. A more detailed investigation would offer better adaptation of this item to the Vietnamese culture. At the bottom of Table 1, it is noted to pay particular attention to this item when the Vietnamese K-JCQ is administered to workers. Finally, all five subscales of the K-JCQ emerging from the Vietnamese sample showed satisfactory internal consistency and good test-retest reliability, considering the small number of items for some subscales (e.g., skill discretion).

For criterion validity, it is interesting to note that the most significant measure of work absences correlated to subscales of the Vietnamese K-JCQ remained the number of times people were absent from the workplace. This can be explained by the fact that almost three-quarters of participants (77%) quoted family reasons for the nature of work absences and just a small proportion indicated that absences were related to the participant's working conditions (less than 15%). In other words, it is not sufficient to hypothesize that work absences are related to the highest level of the K-JCQ subscales (e.g., psychological demands), but it is also

important to consider the nature of absences and the number of times the person is absent from the workplace. In the Vietnamese context, the repetition of absences seems a better indicator context of poor work conditions or poor work satisfaction and would correlate better with the K-JCQ subscales when compared to the work absence (yes or no) or the number of workday absences. On one hand, there was a significant correlation between depression and psychological demand in the workplace and this relationship was supported by several studies (Mausner-Dorsch & Eaton, 2000; Tsutsumi, Kayaba, Theorell, & Siegrist, 2001; Wang & Patten, 2001). On the other hand, and reinforcing the results presented above, when people presented more depressive symptoms, they also mentioned their workplace as having poor working conditions. In summary, these complementary results reveal the effect of working conditions, particularly psychological demands, on the mental health condition of workers (Kitaoka-Higashiguchi, Nakagawa, Morikawa, Ishizaki, Miura, Naruse, *et al.*, 2002; Robiner, 2006). The particular context of the hospital, such as having reduced staff and somewhat poor, stressful working conditions, are recognized in the literature to have a negative effect on the mental health condition of health professionals (Muntaner, Li, Xue, O'Campo, Chung, & Eaton, 2004; Bourbonnais, Brisson, Malenfant, & Vezina, 2005).

This study does have limitations. Firstly, the type of sample chosen for this study might be too specific for the validation of the K-JCQ, though several validation studies have been conducted with health professionals since this population is more likely to develop burnout and other mental health problems. In this study, less than 5% of health personnel presented a higher risk of depression. However, it was not possible to access information related to the number of work absences due to depression in this specific workplace in order to better document the prevalence rate. Secondly, and related to the first limitation, only personal information regarding work absences were used for the validation of the K-JCQ. To date, this limitation is difficult to overcome since objective absence data (e.g., administrative absences) collected by human resources in Vietnamese workplaces is not systematic. This was confirmed by occupational physicians working in Vietnam who reported that administrative data in Vietnam are unreliable. To date, the subjective evaluation of work absences remains the unique measure available and probably the only one that researchers can rely on.

Overall, these results show that the Vietnamese version of the K-JCQ is valid, since this questionnaire respects the conceptual foundations of the original version while expressing some cultural specificity and it demonstrates good psychometric properties (construct validity, internal consistency, test-retest reliability, and criterion validity). Clinical studies on

occupational stress in Vietnam can now be planned. Creating this first valid instrument to measure job strain makes such studies feasible and may benefit Vietnamese workers.

REFERENCES

- BONDE, J. P. E. (2008) Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. *Occupational and Environmental Medicine*, 65(7), 438-445. DOI: 10.1136/oem.2007.038430
- BORTEYROU, X., TRUCHOT, D., & RASCLE, N. (2009) Le stress chez le personnel travaillant en oncologie: une tentative de classification des stratégies de coping. *Psychologie du travail et des organisations INSS*, 16(4), 380-398.
- BOURBONNAIS, R., BRISSON, C., MALENFANT, R., & VEZINA, M. (2005) Health care restructuring, work environment, and health of nurses. *American Journal of Industrial Medicine*, 47(1), 54-64. DOI: 10.1002/ajim.20104
- DE ARAÚJO, T. M., & KARASEK, R. (2008) Validity and reliability of the Job Content Questionnaire in formal and informal jobs in Brazil. *Scandinavian Journal of Work, Environment & Health*, Supplements(6), 52-59.
- DE JONGE, J., REUVERS, M. M. E. N., HOUTMAN, I. L. D., BONGERS, P. M., & KOMPIER, M. A. J. (2000) Linear and nonlinear relations between psychosocial job characteristics, subjective outcomes, and sickness absence: baseline results from SMASH. *Journal of Occupational Health Psychology*, 5(2), 256-268. DOI: 10.1037/1076-8998.5.2.256
- DÉPARTEMENT ÉPIDÉMIOLOGIE EN ENTREPRISE, I. (2006) Stress et risque psychosociaux: concepts et prévention. *INRS Documents pour le Médecin du Travail*, 106, 169-186.
- ELO, A-L., LEPPÄNEN, A., & JAHKOLA, A. (2003) Validity of a single-term measure of stress symptoms. *Scandinavian Journal of Work, Environment & Health*, 29(6), 444-451.
- EUM, K-D., LI, J., JHUN, H-J., PARK, J-T., TAK, S-W., KARASEK, R., & CHO, S-I. (2007) Psychometric properties of the Korean version of the Job Content Questionnaire: data from health care workers. *International Archives of Occupational and Environmental Health*, 80(6), 497-504. DOI: 10.1007/s00420-006-0156-x
- GANSTER, D. C., & PERREWÉ, P. L. (2011) Theories of work stress. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology*. Washington, DC: American Psychological Association. Pp. 37-53.
- HADI, A. A., NAING, N. N., DAUD, A., & NORDIN, R. (2006) Reliability and construct validity of the Malay version of the Job Content Questionnaire (JCQ) among secondary school teachers in Kota Bharu, Kelantan, Malaysia. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 37(6), 1254-1259.
- HOUTMAN, I., KORNITZER, M., DE SMET, P., KOYUNCU, R., DE BACKER, G., PELFRENE, E., ROMON, M., BOULENGUEZ, C., FERRARIO, M., ORIGGI, G., SANS, S., PEREZ, I., WILHELMSSEN, L., ROSENGREN, A., ISACSSON, S-O., & ÖSTERGREN, P-O. (1999) Job stress, absenteeism and coronary heart disease European cooperative study (the JACE study): design of a multicentre prospective study. *The European Journal of Public Health*, 9(1), 52-57. DOI: 10.1093/eurpub/9.1.52
- JOHNSON, J. V., & HALL, E. M. (1988) Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, 78(10), 1336-1342.
- KARASEK, R. (1979) Job demands, job decision latitude, and mental strain: implications for job redesign. *Administrative Science Quarterly*, 24(2), 285-308.

- KARASEK, R. (1985) *Job Content Questionnaire and user's guide*. Lowell: Univer. of Massachusetts, Department of Work Environment.
- KARASEK, R., BRISSON, C., KAWAKAMI, N., HOUTMAN, I., BONGERS, P., & AMICK, B. (1998) The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology*, 3(4), 322-355. DOI: 10.1037/1076-8998.3.4.322
- KARASEK, R., & THEORELL, T. (1990) *Healthy work: stress, productivity, and the reconstruction of working life*. New York: Basic Books.
- KITAOKA-HIGASHIGUCHI, K., NAKAGAWA, H., MORIKAWA, Y., ISHIZAKI, M., MIURA, K., NARUSE, Y., & KIDO, T. (2002) The association between job demand, control and depression in workplace in Japan. *Journal of Occupational Health*, 44, 427-428.
- KOMPIER, M. A. J., & LEVI, L. (1995) *Le stress au travail: causes, effets et prévention*. Livre de Fondation européenne pour l'amélioration des conditions de vie et de travail.
- LEKA, S., GRIFFITHS, A., & COX, T. (2004) *Work organization and stress: systematic problem approaches for employers, managers and trade union representatives*. Protecting workers' health series, no. 3: World Health Organization. Retrieved from <http://whqlibdoc.who.int/pwh/2004/9242590479.pdf>.
- LI, J., YANG, W., LIU, P., XU, Z., & CHO, S. (2004) Psychometric evaluation of the Chinese (mainland) version of Job Content Questionnaire: a study in university hospitals. *Industrial Health*, 42(2), 260-267.
- MAIZURA, H., RETNESWARI, M., MOE, H., HOE, V. C. W., & BULGIBA, A. (2010) Job strain among Malaysian office workers of a multinational company. *Occupational Medicine*, 60(3), 219-224. DOI: 10.1093/occmed/kqq001
- MAUSNER-DORSCH, H., & EATON, W. W. (2000) Psychosocial work environment and depression: epidemiologic assessment of the demand-control model. *American Journal of Public Health*, 90(11), 1765-1770.
- MUNTANER, C., LI, Y., XUE, X., O'CAMPO, P., CHUNG, H. J., & EATON, W. W. (2004) Work organization, area labor-market characteristics, and depression among U.S. nursing home workers: a cross-classified multilevel analysis. *International Journal of Occupational and Environmental Health*, 10(4), 392-400. [Research Support, U.S. Gov't, P.H.S.]
- NAKAO, M. (2010) Work-related stress and psychosomatic medicine. *BioPsychoSocial Medicine*, 4(1), 4.
- NGUYEN, B. D. (2003) Depression diagnosis results among students of a high school in Hanoi. *Vietnam Journal of Psychology*, 7, 57-64.
- NGUYEN, T. T. A. (2013) *Symptoms of depression among patients aged 18-45 years treated in Danang General Hospital from October 2008 to March 2009*. Danang Univer. of Pedagogy.
- NIEUWENHUIJSEN, K., BRUINVELS, D., & FRINGS-DRESEN, M. (2010) Psychosocial work environment and stress-related disorders, a systematic review. *Occupational Medicine*, 60(4), 277-286. DOI: 10.1093/occmed/kqq081
- PELFRENE, E., VLERICK, P., MAK, R. P., DE SMET, P., KORNITZER, M., & DE BACKER, G. (2001) Scale reliability and validity of the Karasek 'Job Demand-Control-Support' model in the Belstress study. *Work & Stress*, 15(4), 297-313.
- ROBINER, W. N. (2006) The mental health professions: workforce supply and demand, issues, and challenges. *Clinical Psychology Review*, 26(5), 600-625. DOI: 10.1016/j.cpr.2006.05.002

- SCHREURS, P. J. G., & TARIS, T. W. (1998) Construct validity of the demand-control model: a double cross-validation approach. *Work & Stress*, 12(1), 66-84. DOI: 10.1080/02678379808256849
- STANSFELD, S., & CANDY, B. (2006) Psychosocial work environment and mental health—a meta-analytic review. *Scandinavian Journal of Work, Environment & Health*, 36(6), 443-462.
- THEORELL, T., PERSKI, A., AKERSTEDT, T., SIGALA, F., AHLBERG-HULTEN, G., SVENSSON, J., & ENEROTH, P. (1988) Changes in job strain in relation to changes in physiological state: a longitudinal study. *Scandinavian Journal of Work, Environment & Health*, 14(3), 189-196. DOI: 10.5271/sjweh.1932
- THOMPSON, B. (1996) AERA editorial policies regarding statistical significance testing: three suggested reforms. *Educational Researcher*, 25(2), 26-30. DOI: 10.2307/1176337
- TSUTSUMI, A., KAYABA, K., THEORELL, T., & SIEGRIST, J. (2001) Association between job stress and depression among Japanese employees threatened by job loss in a comparison between two complementary job-stress models. *Scandinavian Journal of Work, Environment & Health*, 27(2), 146-153.
- VALLERAND, R. J. (1989) Vers une méthodologie de validation transculturelle de questionnaires psychologiques: implications pour la recherche en langue française. *Psychologie Canadienne*, 30(4), 662-680.
- WANG, J., & PATTEN, S. B. (2001) Perceived work stress and major depression in the Canadian employed population, 20-49 years old. *Journal of Occupational Health Psychology*, 6(4), 283-289.
- WORLD HEALTH ORGANIZATION. (2008) Sensibilisation au stress professionnel dans les pays en développement. *Série Protection de la santé des travailleurs*, n° 6. Bibliothèque de l'OMS. Pp. 1-20.

Accepted August 30, 2013.

PSYCHOMETRIC PROPERTIES OF THE ACHIEVEMENT GOAL QUESTIONNAIRE–REVISED ADMINISTERED TO GREEK UNIVERSITY STUDENTS^{1, 2}

MARY APOSTOLOU

University of Ioannina

Summary.—This study examined the generalizability of the Achievement Goal Questionnaire–Revised (AGQ–R) in a different cultural and ethnic context than other published studies and across different levels of specificity in the academic domain. The study assessed the structural validity and reliability of the questionnaire as administered to 105 Greek University students in both a course-specific context and in a more general academic context. Six plausible alternative models with three- and two-factor structures were compared to the four-factor structure of the questionnaire. In the course-specific context administration, the four-factor structure of the AGQ–R was supported, but a three-factor structure model did not provide a worse fit. None of the remaining models provided a good fit. At the more general academic context, the four-factor structure of the questionnaire did not show adequate fit to the data. The same was true for each one of the six alternative models. The internal consistency reliabilities for the factors depicted in the four- and three-factor structure models were very good, ranging from .90 to .98.

The achievement goal construct represents the contemporary perspective in the study of achievement motivation. According to Elliot and Murayama, (2008) “it represents a landmark contribution to the century-long study of competence and motivation” (p. 613). Several measurement instruments have been used for the assessment of achievement goals, most recently the Achievement Goal Questionnaire–Revised (AGQ–R; Elliot & Murayama, 2008). Four achievement goals are measured by the questionnaire: mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance. All four goals are conceptually considered to be cognitive-dynamic goals (Elliot, 1999) focusing on competence. Research evidence suggests that each one of the four achievement goals leads to different cognitive, affective, and behavioral processes and outcomes related to achievement (cf. Elliot, 2005, for a review). So, to ascertain the achievement goals adopted by portions of the population, a reliable and valid research instrument is needed.

¹Address correspondence to Mary Apostolou, Ph.D., Department of Philosophy, Education, and Psychology, Education Section, School of Philosophy, University of Ioannina, Dourouti, 45110 Ioannina, Greece or e-mail (mapostol@uoi.gr).

²The author thanks Mr. George Dimakopoulos, MSc Statistician, for his valuable assistance in the statistical analysis of the research data.

A body of research, using various instruments for the measurement of achievement goals, has examined the generalizability of achievement goals across different levels of specificity in the academic domain and across different cultures. Nevertheless, relevant research data concerning the generalizability of the AGQ-R with non-American samples are very rare. There is also a lack of empirical literature related to the generalizability of the AGQ-R when used with a Greek population. Thus, the psychometric properties of the AGQ-R in a Greek student sample were examined after translation into the Greek language and across different levels of specificity in the academic domain. Additionally, a comparative examination was made between the four-factor structure of the AGQ-R and six theoretically possible alternative models with three- and two-factor structures.

The study of achievement motivation was enriched in the recent years by the achievement goal construct (Dweck & Leggett, 1988), which emphasizes the concept of competence. Essentially, when people are achievement-motivated, they work towards the fulfillment of their need for competence (for reviews see Elliot, 1999; Elliot, McGregor, & Thrash, 2002; Dweck & Molden, 2005; Harackiewicz & Linnenbrink, 2005). The work on this construct emerged from classroom research within educational psychology, but it was extended to other types of achievement, such as sport and physical education settings and the work domain, and it has been proposed for use in clinical psychology and counseling (Dykman, 1998; Conroy & Elliot, 2004; Barron, Baranick, & Finney, 2006; Wang, Biddle, & Elliot, 2007).

The achievement-goal construct has been elaborated conceptually. Dweck and colleagues suggested a dichotomous model that distinguished between two different achievement goals: performance and learning (Dweck, 1986; Dweck & Leggett, 1988; Elliot & Dweck, 1988). Several leading researchers have used different terminology for these two classes of goals (Nicholls, 1984; Ames & Archer, 1988; Maehr & Midgley, 1991; Elliot, 1999). According to Ames (1992) there exists a conceptual differentiation between these two groups of constructs that could be unified on a theoretical basis under the labels of "mastery" and "performance" goals. When individuals pursue a mastery goal, they focus on self-referenced criteria and they try to develop new skills, understand their work, enrich their competence and gain a sense of personal mastery. When they adopt a performance goal, they concentrate on normative competence criteria and they try to do better than others.

Initially, research findings (Dweck, 1986; Dweck & Leggett, 1988; Elliot & Dweck, 1988) indicated that mastery goals were connected with a number of positive cognitive, affective, and behavioral achievement-relevant processes and outcomes associated with an adaptive, mastery-oriented response pattern. They also indicated that performance goals were associ-

ated with a constellation of negative cognitive, affective, and behavioral achievement-related processes and outcomes connected with a maladaptive, helpless-oriented response pattern.

More recently, researchers have questioned the belief in the “good” or “bad” nature of the mastery and performance goals respectively (see Elliot, 2005, for review). Elliot and co-researchers (Elliot & Harackiewicz, 1996; Elliot & Church, 1997) argued that the varying findings, mainly concerning performance goals, might be attributed to the type of measures used which were taking into consideration dichotomous models and were considering only the definition (mastery-performance distinction) dimension of the competence construct. They recognized Dweck's (1986) distinction of approach-avoidance in the definition of a performance goal (Dweck & Bempechat, 1983), and specifically her claim that students adopting a performance goal orientation seek to “gain favorable judgments of their competence or avoid negative judgments of their competence” (p. 1040). Consequently, they proposed that the construct of competence should be differentiated along a second dimension related to the concept of valence, which would include the approach-avoidance distinction. This distinction first appeared in the “big” historic motivation theory of Lewin (1935), who used it in his Field Theory to conceptualize positive and negative valences. Subsequently, the work of McClelland, Atkinson, Clark, and Lowell (1953) on approach and avoidance motives helped to make this concept more acknowledged and well established. Nevertheless, we must point out that similar conceptual definitions have been used in the past, starting with the ancient Greek philosophers Democritus (460–370 B.C.), Aristippus (435–355 B.C.), and Epicurus (341–270 B.C.) (for review see Elliot & Covington, 2001). Hence, there were two dimensions on which the concept of competence should be differentiated: the definition dimension relating to the way a person evaluates individual performance, and the valence dimension relating to a person's tendency to focus either on approaching or on avoiding (see Elliot, 1997a, 1997b, 1999; Elliot & Covington, 2001; Elliot & Thrash, 2001; Elliot, 2006; for reviews on the approach-avoidance distinction).

According to this point of view, competence may be defined in terms of each one of three evaluative referents, which compose the mastery definition of the competence construct: (1) the absolute standard of what the task requires, (2) the intrapersonal standard of past performance, (3) the maximum potential for attainment. Competence may also be defined in terms of an interpersonal normative standard relating to other persons' performance, which is the performance definition of the competence construct (Elliot & McGregor, 2001). Competence may be valenced depending on the personal evaluative focus. If that focus allows the person to evaluate his performance or ability as positive and potentially successful,

it evokes approaching forms of behavior. If it is focused on a negative and potentially unsuccessful possibility, it evokes avoiding forms of behavior (Elliot & McGregor, 2001).

The combination of the definition dimension (mastery-performance distinction) and the valence dimension (approach-avoidance distinction) of the competence construct resulted initially in a trichotomous achievement goal framework (Elliot & Harackiewicz, 1996; Elliot & Church, 1997) in which the performance goal construct was bifurcated in terms of the valence dimension (the approach-avoidance distinction) and the mastery goal construct was left intact. Later, Elliot and his colleague (Elliot & McGregor, 2001) developed the achievement goal framework by partitioning the mastery goal construct into two valenced components: a mastery-approach goal and a mastery-avoidance goal. The result was the formulation of a 2×2 achievement goal framework consisting of four types of achievement goals (Elliot & McGregor, 2001; Elliot, 2008): *mastery-approach goals*, in which an attempt is made to approach task-based or intrapersonal competence, as, for example, in the case where a person makes an effort with the purpose of surpassing past individual achievements or tries to perfectly master the material presented in a class based on a personal criterion of accomplishment; *mastery-avoidance goals*, in which an attempt is made to avoid task-based or intrapersonal incompetence, as, for example, when a person tries to avoid learning less than it would be possible according to a personal standard of attainment or makes an effort not to do worse compared to past achievements; *performance-approach goals*, in which an effort is made to approach interpersonal competence, as, for example, when a person engages in a task with the purpose of doing better than his classmates; *performance-avoidance goals*, in which an effort is made to avoid interpersonal incompetence, as in the case where a person becomes involved in a task with the purpose of avoiding being proved inferior to her classmates.

Generalizability of AGQ-R Across Different Academic Contexts

One most recent measurement instrument designed to assess the four-dimensional conceptualization of the achievement goal construct is the Achievement Goal Questionnaire-Revised (AGQ-R) (Elliot & Murayama, 2008). It was designed to assess achievement goals within a course-specific context. However, the issue related to the extent to which the 2×2 structure may be generalized to a more general academic context is still unsettled. Responses to a situationally specific academic achievement goal questionnaire may reflect the influence of situational characteristics and salient features of the specific situation and the given context, or they

may reflect the effect of typical and rather stable attitudes and beliefs in the academic domain (Finney, Pieper, & Barron, 2004).

The question of whether goal orientation is a trait or a state has been the subject of studies by several researchers. In their review, Harackiewicz, Barron, and Elliot (1998) explained how the construct has been operationally defined and researched within different levels of specificity. Commenting on the subject of what motivates students to act in specific classroom situations, Boekaerts (2001) addressed the question of whether goal setting and goal striving is influenced by general, domain-specific, and situation-specific motivation (see Volet & Järvelä, 2001, for research on motivation in learning contexts). This issue is regarded by DeShon and Gillepsie (2005), in their review of the subject, as "one of the most vexing problems in the goal orientation literature.... Is it stable over situations and domains, or is it situation and domain specific? Is it stable over time within the same situation (e.g., a team meeting)?" (p. 1115). They pointed out that, although a substantial body of research has been carried out, we do not know yet whether goal orientation reflects a behavioral disposition, which invariably is made evident in a range of situations or a temporary way of behaving. In their empirical review of the relevant literature, Payne, Youngcourt, and Beaubien (2007) posed similar questions concerning the stability of trait goal orientation over time and the extent to which trait (individual dispositional difference) and state (way of responding specific to a given task and context) goal orientations are interrelated. In line with this speculation, Pintrich (2000) argued that an important future research topic concerning achievement goals should be to examine how stable these goals are over time, contexts, and domain dimensions. Vandewalle (1997) argued that achievement goal measures should contain items operationalized at a midlevel of specificity, such as academic, work, and athletic fields, which constitute major life domains. He stressed the inadequacy of past measurement instruments in tapping goal orientations in specific areas, because those instruments were mainly global or specific to a context. He also emphasized the need for better evaluation of construct validity evidence of achievement goal measures through the use of confirmatory factor analyses, which provide additional knowledge and evidence from exploratory factor analyses, which are the usual type of analysis. This point was also made by Ingles, Garcia-Fernandez, Castejon, Valle, Delgado, & Marzo (2009). The additional knowledge obtained could be helpful in establishing the broad application of the achievement goal framework.

To our knowledge, there is lack of research examining the extent to which the AGQ-R can be generalized within different levels of specificity in the academic domain. Research using its predecessor, the Achievement Goal Questionnaire (AGQ) (Elliot & McGregor, 2001), supported its

four-factor structure not only in situation-specific or academic contexts, but in more general or in other than academic achievement contexts as well (Elliot & Reis, 2003; Van Yperen, 2003; Conroy & Elliot, 2004; Finney, *et al.*, 2004; Wang, *et al.*, 2007; Stevenson & Lochbaum, 2008).

Generalizability of AGQ-R Across Different Cultures

Questions about generalizability also apply to assessment instruments validated within specific cultural and ethnic groups and to whether their factor structure remains invariable within other countries and diverse ethnic and cultural contexts (Urdu, 1997; Midgley, Kaplan, & Middleton, 2001; Kaplan, Middleton, Urdu, & Midgley, 2002; Yu & Wolters, 2002; Urdu, 2004; Harackiewicz & Linnenbrink, 2005). For example, Graham and Hudley (2005) observed that experiences with discrimination and immigration, being the target of racial stereotypes and having a certain racial and ethnic identity, can all affect achievement motivation and shape achievement strivings and the pursuit of competence.

Since the achievement goal perspective is regarded as a major theoretical approach to the study of achievement motivation, it could be claimed that the previously mentioned variables might influence achievement goals as well. So, one might assume that samples of respondents with different backgrounds in ethnic, racial, and cultural identity could yield different structures of achievement goals. In line with this reasoning, Kaplan and Maehr (2007) argued that certain variables (e.g., socio-cultural background, ethnicity, various subject matters) may have an effect on the structure and meaning of the achievement goal construct. The authors suggested that "goal orientation theory ... would benefit ... from a cross-cultural approach to motivation in different cultural groups, with research that focuses on the socio-cultural meaning of action in achievement settings within cultural as well as other types of groups" (p. 149). Because of the possible influence of the cultural context on the structure and conceptualization of achievement goals, McInerney, Roche, McInerney, and Marsh (1997) emphasized that it is necessary to examine an entire scale validated in one cultural context, to assess its applicability in different cultural settings. Currently, there exists a research body examining cross-cultural differences in the nature of achievement goals (Ommundsen, 2001; Dupeyrat & Mariné, 2005; Van Yperen, 2006; Lau & Lee, 2008; David, 2009; Ingles, *et al.*, 2009; Murayama, Zhou, & Nesbit, 2009; Alkharusi & Aldhafri, 2010). The comparison of achievement goal factor structures in different racial and ethnic groups has also attracted the interest of researchers (Campbell, Barry, Joe, & Finney, 2008).

The existence of a differential link between achievement goals and various positive or negative achievement-related processes and outcomes

has been supported by many studies (for review, see Elliot, 2005). One implication is that it may be possible to influence the development of goal orientations which relate to positive rather than negative cognitive, affective and behavioral achievement-related variables and outcomes (Kaplan & Maehr, 2007; Campbell, *et al.*, 2008). For such potential interventions to be studied, a valid and reliable achievement goal assessment instrument is needed which works equally well within different cultural and ethnic groups. Currently, there is lack of validity and reliability evidence for the AGQ-R scores, which makes it difficult to assess the generalizability of the 2×2 achievement goal framework and, as a consequence, to make cross-cultural comparisons of research results. It is still unknown whether people from different ethnic and cultural backgrounds have the same conceptualization for the achievement goal construct and whether measures representing the construct function in an equivalent way for all groups (Campbell, *et al.*, 2008).

In the present study, the core research question was related to the generalizability of the AGQ-R, and one purpose was to provide structural validity and reliability evidence based on the scores of the AGQ-R when administered to a sample of Greek University students in both a course-specific academic context and in a more general academic context. An additional purpose was to compare the four-factor structure of the AGQ-R model to the structure of four three-factor and two two-factor, alternative models (Elliot & Murayama, 2008) within a course-specific and a more general academic context. Elliot and Murayama (2008), using an American sample, have found that "alternative factor models could not explain the data better than the four-factor model" (p. 620) in a course-specific academic context.

The final purpose of the study was to examine the cross-cultural generalizability of the AGQ-R by providing structural validity and reliability evidence of its scores when administered to a sample of Greek University students (a) in a course-specific academic context and (b) in a more general academic context. In addition, the questionnaire's four-factor structure was compared to four alternative models (three-factor and two two-factor) (Elliot & Murayama, 2008).

The construct of achievement goals, assessed with measures other than the AGQ-R, has been investigated with Greek samples by several researchers (Leontari & Gialamas, 2002; Sideridis, 2005, 2006a, 2006b; Barkoukis, Ntoumanis, & Nikitaras, 2007; Barkoukis, Thøgersen-Ntoumani, Ntoumanis, & Nikitaras, 2007; Kouli & Papaioannou, 2009). The AGQ-R was also employed with non-American samples by Alkharusi and Aldhafri (2010), Abd-El-Fattah and Al-Nabhani (2012), and by Carette, Anseel, and Van Yperen (2011) who used Omani and Belgian samples, respec-

tively. This study is the first time the AGQ-R has been used simultaneously within two different levels of specificity in the academic domain and the first time it has been used in Greece.

METHOD

Participants

The research data were collected from a sample of 105 Greek students (90 women, 15 men) enrolled in a Department of Philosophy-Education-Psychology in a public Greek university. Their ethnic identity was Greek and their mean age was 19.7 yr. ($SD = 0.7$). The unequal number of male and female students in the research sample was unavoidable, since the enrollment Department belongs to the School of Philosophy, which in the Greek Universities traditionally attracts a larger proportion of female students.

Measure

The Achievement Goal Questionnaire-Revised (AGQ-R: Elliot & Murayama, 2008) is the result of an evaluative critique of the Achievement Goal Questionnaire (AGQ: Elliot & McGregor, 2001), which provided the conceptual basis for its structure. It is a 12-item questionnaire with four subscales, each consisting of three items and assessing each of the four achievement goals included in the 2×2 achievement goal framework: mastery-approach goals (e.g., "my aim is to completely master the material presented in this class"), mastery-avoidance goals (e.g., "my aim is to avoid learning less than I possibly could"), performance-approach goals (e.g., "my aim is to perform well relative to other students"), performance-avoidance goals (e.g., "my aim is to avoid doing worse than other students").

The intact AGQ-R was administered in a course-specific context. For the administration in the more general academic context, some items were revised and the depicted achievement goals operationalized in terms of this context. In particular, the words "class" and "course" were replaced in Items 1, 2, and 5 with the wording "courses this semester"; so, these items were rewritten as follows: Item 1: "my aim is to completely master the material presented in the courses this semester"; Item 2: "I am striving to understand the content of the courses this semester as thoroughly as possible"; Item 5: "I am striving to avoid an incomplete understanding of the courses' material this semester." Additionally, one of the questionnaire's completion directions drew participants' attention to the wording "*courses this semester*." The 12 items were presented in the following random order: 1, 8, 3, 7, 4, 12, 2, 9, 6, 11, 5, 10. The questionnaire was given in the form of an adjective rating scale with anchors 1: Strongly disagree and 5: Strongly agree, with a higher score meaning a stronger endorsement of the achievement goal. Kaplan and Maehr (2007) observed that it is cur-

rently a very common research procedure to use adjective rating scales for the assessment of achievement goals.

For the purpose of establishing content validity, the AGQ-R was translated into Greek by a process of back translation (Hambleton & Kanjee, 1995) made by two native Greek speakers with proficient knowledge of the English language. Initially, the AGQ-R was translated into Greek by the author and, when completed, the Greek translation was back-translated into English by the author's colleague. Next, a comparison was made between the back translation and the original questionnaire and some corrections considered appropriate by the two translators were made to the final Greek translation.

Procedure

The students who participated in the research project were attending an introductory Psychology class and volunteered to participate at the author's request. The students completed the AGQ-R at the end of a regular three-hour class period, two weeks before the examination. They were attending lectures for a semester course, and evaluation was based on examination performance attainment. With regard to the counterbalancing of the order, half of the sample population first answered the questionnaire form as it was applied in the course-specific context and after that they completed the form as adapted to a more general academic context. The other half answered the questionnaires the other way round. The questionnaires were distributed by the author, assisted by two postgraduate students.

Statistical Analysis

The structural validity evidence of the scores was based on confirmatory factor analysis (CFA) and on the Pearson product-moment correlations between the scores on the AGQ-R subscales. The internal consistency reliability of the AGQ-R scores was based on estimation of CFA reliabilities.

RESULTS

Descriptive Statistics and Reliability Analyses

With reference to CFA-based reliabilities, the acceptable cutoff value is considered to be $>.60$ (Bagozzi & Yi, 1988). The descriptive statistics (means, standard deviations, observed minimum, and maximum values) and CFA-based reliabilities for each one of the four AGQ-R measures in both administrations are presented in Table 1.

In the specific context, the students endorsed each of the four achievement goals and the mean score for each subscale was above its midpoint. Students scored highest on the mastery-approach measure, followed by the mastery-avoidance, performance-avoidance and performance-approach

TABLE 1
DESCRIPTIVE STATISTICS AND RELIABILITY EVIDENCE OF SCORES ON THE AGQ-R ($N = 105$)

Context	Variable	M	SD	Min.	Max.	Reliability	
						CFA-based	Cronbach's α
Course-specific context	Mastery-approach	4.03	0.55	2.33	5.00	.90	.51
	Mastery-avoidance	4.02	0.69	2.00	5.00	.94	.62
	Performance-approach	3.04	0.90	1.00	5.00	.98	.85
	Performance-avoidance	3.26	0.91	1.00	5.00	.98	.84
General academic context	Mastery-approach	3.97	0.56	2.00	5.00	.92	.50
	Mastery-avoidance	4.03	0.59	2.00	5.00	.86	.74
	Performance-approach	2.99	0.76	1.00	4.33	.96	.87
	Performance-avoidance	3.16	0.83	1.00	4.67	.97	.87

measures. With reference to CFA-based reliabilities for the four measures, the scores' reliability estimate for the items depicting performance-approach was .98, performance-avoidance was .98, mastery-approach was .90, and mastery-avoidance was .94. Cronbach's α coefficients were .85, .84, .51, and .62, respectively.

In the general context, the students' highest mean score was on the mastery-avoidance measure, which was followed by the mastery-approach, performance-avoidance and performance-approach measures. The CFA-based scores' reliability estimate for the items referring to the performance-approach was .96, to the performance-avoidance .97, to the mastery-approach .86, while the mastery-avoidance measure had a reliability value of .92. Cronbach's α coefficients were .87, .87, .50, and .74, respectively.

CFA Validity Evidence Based on the Internal Structure of Achievement Goal Measures

Confirmatory factor analysis (CFA) helps researchers evaluate models grounded on a theoretical basis for the structure of the relationships in a number of latent variables and, through a direct comparison of several alternative factor structures, to assess which structure offers the best fit to the data (Murayama, *et al.*, 2009). A series of CFAs using the Mplus 6.0 software program (Muthen, 1989a; 1989b) were conducted on the achievement goal items of the AGQ-R and a set of six alternative models to compare their fit to the data: (a) a trichotomous Model A, consisting of a performance-approach factor, a performance-avoidance factor, and a combined mastery-approach and mastery-avoidance factor, (b) a trichotomous Model B, consisting of a mastery-approach factor, a mastery-avoidance factor, and a combined performance-approach and performance-avoidance factor, (c) a trichotomous Model C, consisting of a mastery-approach

factor, a performance-approach factor, and a combined mastery-avoidance and performance-avoidance factor, (d) a trichotomous Model D, consisting of a mastery-avoidance factor, a performance-avoidance factor, and a combined mastery-approach and performance-approach factor, (e) a mastery-performance dichotomous Model E, consisting of a combined mastery-approach and mastery-avoidance factor, and a combined performance-approach and performance-avoidance factor, (f) an approach-avoidance dichotomous Model F, consisting of a combined mastery-approach and performance-approach factor, and a combined mastery-avoidance and performance-avoidance factor. The six alternative models were compared to the four-factor structure depicted in the AGQ-R; they were more parsimonious and corresponded to six theoretically plausible competing combinations of the definition and valence dimensions of the achievement goal construct. It should be noted that the empirical evidence has not supported the factor structures depicted in all these models. But since the AGQ-R is a most recent four-dimensional conceptualization of the achievement goal construct, it is worthwhile to examine all the theoretically possible factor structures depicted in the six alternative models. In doing this comparison it was desired, firstly, to clarify whether any of the six models was a poorer fit to the data than the four-factor structure model. If that was the case, this study would be in a position, considering any additional research results, to champion a less complex model than the AGQ-R (Campbell, *et al.*, 2008). Also, this study wanted to ascertain whether the recently added, to the achievement-goal construct conceptualization, mastery-avoidance goal is a distinct form of regulation that can be empirically distinguished from the mastery-approach goal (difference in terms of the valence of competence) and the performance-avoidance goal (difference in terms of the definition of competence) (Elliot & McGregor, 2001; Elliot & Murayama, 2008), or be collapsed into an undivided mastery factor (approach-avoidance) and an undivided avoidance factor (mastery-performance), respectively.

According to the recommendations of Hoyle and Panter (1995), the analyses were conducted on covariance matrices and the solutions were generated on the basis of maximum-likelihood mean adjusted. Following the suggestions made by Hu and Bentler (1995, 1998) and Hoyle and Panter (1995), multiple indexes were employed to assess the models' fit to the data: a chi-square statistic divided by its degrees of freedom (χ^2/df), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) index with confidence intervals. Because the intention was to compare several models, the Akaike information criterion (AIC) was used as the comparison statistic for nonhierarchical models (Grant & Dweck, 2003; Elliot & Murayama, 2008) along with the sample-size adjusted BIC information criterion (SABIC) (Sclove, 1987; Tofighi & Enders, 2007; Enders & Tofighi, 2008).

The following criteria were used to evaluate the adequacy of the models' fit to the data: for the χ^2/df index a value equal or under 2.0 ($\chi^2/df \leq 2.0$) (Jöreskog, 1969), for the CFI index a value equal or above .90 ($CFI \geq .90$) (Hu & Bentler, 1995), and for the TLI index a value equal or above .90 ($TLI \geq .90$) (Hu & Bentler, 1998). With reference to the RMSEA index, Browne and Cudeck (1993) suggested that a value up to .05 indicates a close fit of the model to the data, values greater than .05 but less than .08 indicate a relatively good fit, values less than .10 indicate a fair fit, while values greater than .10 indicate an unacceptable fit. As far as the Akaike information criterion and the SABIC information criterion are concerned, an increasingly better fit is indicated by increasingly lower values.

CFA of the hypothesized 2×2 model in the course-specific context.—Initially, CFA was used to evaluate the adequacy of the hypothesized 2×2 model's fit to the data obtained within the course-specific context. The results were: $\chi^2(48, N = 105) = 68.69, p < .027, \chi^2/df = 1.43, CFI = .94, TLI = .92, RMSEA = .065$ (95%CI = .023, .098). The values obtained for the three (χ^2/df , CFI, IFI) fit indexes reached the recommended criteria for a good fit to the data. The RMSEA slightly exceeded the level of .05, but a criterion value ranging between .05 and .08 might be considered as reflecting a relatively good fit (Browne & Cudeck, 1993).

The examination of standardized factor loadings indicated that all loadings were statistically significant and that the items consisting the mastery-approach subscale ranged from .36 to .71, the mastery-avoidance subscale from .40 to .76, the performance-approach subscale from .70 to .89, and the performance-avoidance subscale from .80 to .86. Item 2 ("I am striving to understand the content of this course as thoroughly as possible") in the mastery-approach subscale, had the lowest loading (<.40).

The additional CFAs which examined the fit of each one of the six alternative measurement models (Elliot & Murayama, 2008) and compared their fit to that of the hypothesized 2×2 model indicated that the values obtained for the four fit measures referring to the trichotomous Model A were: $\chi^2(51, N = 105) = 73.34, p < .021, \chi^2/df = 1.43, CFI = .94, TLI = .92, RMSEA = .066$ (95%CI = .026, .097). Model A had a very similar fit to that of the hypothesized 2×2 structure model. Three of the four indices (χ^2/df , CFI, TLI) gave values within the recommended acceptable levels, and the value of the RMSEA index was slightly above the threshold of .05, which reflects a relatively good fit.

With reference to the 2×2 structure model and Model A, the values of the Akaike information criterion (AIC) and the SABIC information criterion (SABIC) were so faintly slightly different, indicating no considerable difference in the fit of the two models. The Satorra-Bentler scaled χ^2 difference test (Satorra, 2000) indicated that Model A did not provide a worse fit than the hypothesized 2×2 structure model.

CFA-based reliabilities for the scores for each Model A measure were: for the combined mastery-approach and mastery-avoidance, .93; for the performance-approach, .98; and for the performance-avoidance, .98. The findings resulting from the examination of the standardized factor loadings for Model A indicated that for the items representing the undifferentiated mastery measure the loadings ranged from .33 to .72 (in this case Items 2, 3, and 5 had loadings < .40); the items of the performance-approach measure ranged from .70 to .92; and the items of the performance-avoidance measure ranged from .81 to .88.

None of the remaining alternative models evidenced a satisfactory fit to the data. Results showed that χ^2/df varied between 1.70 and 2.60, CFI between .79 and .91, TLI between .74 and .88 and RMSEA between .082 and .124.

The results from the CFAs and the Satorra-Bentler scaled chi-square difference test are shown in Table 2.

TABLE 2
FIT INDEX VALUES FOR THE 2 × 2 MODEL AND ALTERNATIVE MODELS
IN THE COURSE-SPECIFIC CONTEXT

Model	<i>df</i>	χ^2	χ^2 /df	CFI	TLI	RMSEA	90% CI		AIC	SABIC
Four-factor model	48	68.69*	1.43	.94	.92	.065	.023	.098	2874.1	2851.8
Trichotomous A	51	73.34*	1.43	.94	.92	.066	.026	.097	2874.8	2854.1
Trichotomous B	51	85.56‡	1.67	.91	.88	.082	.050	.111	2889.7	2968.1
Trichotomous C	51	116.45‡	2.28	.83	.79	.112	.085	.139	2930.1	2909.3
Trichotomous D	51	102.68‡	2.01	.87	.83	.100	.072	.127	2908.8	2888.1
Dichotomous E	53	90.14‡	1.70	.90	.88	.083	.052	.112	2892.6	2872.1
Dichotomous F	53	136.35‡	2.57	.79	.74	.124	.099	.150	2949.8	2930.1
Model	Satorra-Bentler scaled χ^2 difference test									
	d0-d1	TRd	<i>p</i>							
Four-factor model										
Trichotomous A	3	4.60	.20							
Trichotomous B	3	16.10†	<.01							
Trichotomous C	3	40.11†	<.01							
Trichotomous D	3	43.70†	<.01							
Dichotomous E	5	19.84†	<.01							
Dichotomous F	5	65.43†	<.01							

* $p < .05$, † $p < .01$, ‡ $p < .001$.

Correlations Between Scores on Achievement Goal Measures

With reference to the AGQ-R measures, the four measures were positively and significantly correlated. However, there was no statistically

TABLE 3
CORRELATIONS BETWEEN SCORES ON THE ACHIEVEMENT GOAL MEASURES

Model	Variable	Mastery-approach	Mastery-avoidance	Performance-approach
AGQ-R	Mastery-approach			
	Mastery-avoidance	.53†		
	Performance-approach	.25*	.11	
	Performance-avoidance	.33†	.31†	.76†
	Variable	Mastery	Performance-approach	Performance-avoidance
Trichotomous Model A	Mastery			
	Performance-approach	.21*		
	Performance-avoidance	.36†	.76†	

$p < .05$ (2-tailed), $\dagger p < .01$ (2-tailed).

significant correlation between the performance-approach and the mastery-avoidance measures (Table 3). The two measures sharing the mastery definition of competence (mastery-approach and mastery-avoidance) and the two measures sharing the performance definition of competence (performance-approach and performance-avoidance) were those with the strongest positive correlations.

Two observations can be made with regard to these correlational findings, which were similar to those in the Elliot and Murayama study (2008). The first is that stronger correlations were observed between measures sharing in their items a common definition dimension, rather than a common valence dimension. The second is that the strongest correlation was found between performance-approach and performance-avoidance measures, suggesting students who rated performance-approach goals more strongly were likely to rate performance-avoidance goals similarly. The weakest, non-significant correlation was found between mastery-avoidance and performance-approach measures and can be explained on the grounds that neither of these measures shares in its items the same definition and valence dimensions of competence.

With reference to the correlations between scores on the trichotomous Model A measures, results (Table 3) indicated that the three measures were positively correlated at a statistically significant level. The strongest correlation was between performance-approach and performance-avoidance measures.

CFA of Hypothesized 2 × 2 Model in General Academic Context

The results indicated that there was not a good fit of the hypothesized 2 × 2 structure model to the data. None of the fit indexes reached the

acceptable criteria of an adequate fit: $\chi^2(48, N = 105) = 135.72, p < .0001, \chi^2/df = 2.82, CFI = .85, TLI = .80, RMSEA = .132$ (Table 4).

Similarly to the previous outcomes of the CFAs conducted to examine the factor structure of each one of the six alternative achievement goal models (trichotomous Models A, B, C, and D, and dichotomous Models E and F), none of these models offered a good fit to the data. For each one, all four of the fit indexes were inadequate. Synoptically, for the six alternative models, the results showed that: χ^2/df varied between 2.80 and 3.25, CFI between .80 and .84, TLI between .75 and .80, and RMSEA between .112 and .147 (Table 4).

TABLE 4
FIT INDEX VALUES FOR THE 2 × 2 MODEL AND ALTERNATIVE MODELS IN THE MORE
GENERAL ACADEMIC CONTEXT

Model	df	χ^2	χ^2/df	CFI	TLI	RMSEA	90% CI	AIC	SABIC
Four-factor model	48	135.72‡	2.82	.85	.80	.132	.106 .158	2514.9	2493.8
Trichotomous A	51	146.21‡	2.86	.84	.79	.133	.108 .159	2521.5	2501.8
Trichotomous B	51	142.81‡	2.80	.84	.80	.131	.106 .157	2519.6	2499.9
Trichotomous C	51	156.45†	3.06	.83	.79	.112	.085 .139	2632.6	2677.3
Trichotomous D	51	165.98‡	3.25	.80	.75	.147	.122 .172	2548.1	2528.4
Dichotomous A	53	152.34‡	2.87	.83	.79	.134	.109 .159	2527.1	2508.4
Dichotomous B	53	156.12‡	2.94	.82	.75	.127	.099 .161	2543.1	2576.9

† $p < .01$, ‡ $p < .001$.

DISCUSSION

This study assessed the structural validity and reliability evidence for the Achievement Goal Questionnaire–Revised with a sample of Greek University students in a course-specific and a more general academic context. The fit of six theoretically plausible alternative models was also investigated to compare their fit to that of the four-factor AGQ–R model. The questionnaire was translated, for the first time, into the Greek language.

With regard to the evaluation of the internal consistency of scores of the AGQ–R, all four measures, in both contexts, had high reliabilities. The lowest reliability was obtained for the responses to the mastery-approach measure in the more general academic context administration (.86). As far as the internal consistencies of the measures of trichotomous Model A were concerned, all three measures had very high levels of internal consistency above .90.

When referring to the course-specific context, two achievement goal measurement models, the four-factor structure model, and trichotomous

Model A provided a quite similar good fit to the data. In both cases, the value of the RMSEA index indicated a relatively good fit to the data, and the values of the other fit indexes indicated a close model-data fit. One explanation of the quite similar fit of the two models may be related to the nature of the reciprocating value of the adoption by the students of mastery-approach and mastery-avoidance goals. It is possible that the students' orientation toward the accomplishment of each one of these goals does not have, in their way of thinking, an extended time perspective. This may happen because they suspect that the social environment in Greece, inside and outside the University, is comprised of and relies mainly on, competitive assessment and selection systems (e.g., personnel assessment criteria), which are not connected with competence perceptions based on self-referential criteria. As a result, the intrapersonal pursuit of approaching or avoiding mastery achievement (as it is the case with the four-factor model) may have the same meaning attached to the intrapersonal pursuit of overall mastery achievement (as it is the case with the trichotomous Model A). In such a case, it does not matter whether someone is getting academically better or worse according to self-referenced standards; mastery-approach and mastery-avoidance goals may be considered equivalent. Even in the case where the students differentiate between mastery-approach and mastery-avoidance goals, they may think in terms of an overall mastery achievement goal.

Results indicated that, although the trichotomous Model A did not fit the data statistically significantly worse, it also did not fit better than the four-factor structure model. Of course, it is a more parsimonious model explaining the data in a simpler way. On the other hand, the four-factor model is a more sophisticated model and theoretically consistent with the four-dimensional conceptualization of the achievement goal construct, in which the mastery goal construct has been extended into mastery-approach and mastery-avoidance, which was not the case with the trichotomous Model A. In this model, a performance-approach factor, a performance-avoidance factor, and a combined mastery factor, which includes mastery-approach and mastery-avoidance items, are depicted. Furthermore, the examination of the standardized factor loadings indicated that in the four-factor model only one item in the mastery-approach measure had a loading under .40. In the trichotomous Model A, three items in the mastery measure had loadings under .40. On the grounds of the above-mentioned arguments, the four-factor structure model as depicted in the AGQ-R is the better model, despite the fact that the trichotomous Model A is more parsimonious.

The findings did not support the use of the AGQ-R in the more general academic context, which was not the intention of its designers in the

first place. Thus, it appears the questionnaire cannot be adequately used at this level of specificity; the same applies to the six alternative achievement goal measurement models. So, with reference to whether achievement goals remain stable across various contexts, e.g., within different levels of specificity in the academic domain, or when situationally specific factors exert an influence, the results supported the dynamic nature of achievement goals.

The previously mentioned finding is related to the issue of whether it is allowable for researchers to change a questionnaire item's domain or scale before making sure that this is a justifiable procedure as far as its psychometric properties are concerned. In our study, the adaptation of the questionnaire's items for a more general academic level of specificity was detrimental to its construct validity. Construct validity is an integrated judgment of evaluation referring to the extent to which theoretical ideas and hypotheses and empirical evidence collected from test scores sustain adequate and appropriate inferences and actions (Messick, 1989). The findings did not provide evidence supporting the construct validity of the AGQ-R when shifted to a more general level of specificity in the academic domain.

Pintrich (2000) discussed the issue of "trait versus state-like nature of motivational constructs" (p. 101) and argued that the assumed nature of goals as cognitive representations means that they are affected by factors related to the context as well as to the acting person. In a chapter referring to the new research tendencies on motivation in learning contexts, Volet (2001) maintained that in the interplay between contextual and situational characteristics and general motivational beliefs and values, more research is needed to obtain a clearer picture of the variables underlying the stable or dynamic nature of motivation. Arguing along similar lines, Dermitzaki and Efklides (2001) emphasized that, to understand the nature of student motivation in learning settings, it is important to consider both the context in which the achievement task is situated as well as the personal general or stable motivational tendencies and personality characteristics of the acting person.

The existence of strong correlations between performance-approach and performance-avoidance goals has been a matter of interest. Roeser (2004) questioned "whether these goals denote separate psychological realities or not" (p. 283). He argued that, linguistically and psychologically, the same meaning is attached to the goal of being superior and to the goal of being inferior relative to others, because the first implicitly invokes the second. Another approach concerning this topic is related to the role played by experiences of failure in the potential changing of performance-approach to performance-avoidance goals (Elliot & Harackiewicz, 1996).

They wondered "...whether individuals can maintain a performance-approach goal in the face of failure or whether failure would inevitably elicit a performance-avoidance orientation" (p. 472). They suggested that there are patterns in failure, like recurrent failure episodes, which convey to the person experiencing them the negative message of having low competence in a task or in comparison with others. It is quite possible that such negative feedback may lead to the adoption of a performance-avoidance orientation. On the other hand, they argued that the acceptance of a performance-approach goal in the face of failure might evoke directional goals which "may affect reasoning through reliance on a biased set of cognitive processes—that is, strategies for accessing, constructing and evaluating beliefs" (Kunda, 1990; p. 480). It may be possible that the biased accessing of related structures of knowledge constitutes the foundation and the supporting frame for the adoption of performance-approach goals in the face of failure or difficulties. Referring to a statistically significant, strong correlation ($r = .68$) between performance-approach and performance-avoidance goals found in their study, Elliot and Murayama (2008) stressed the importance of exploring possible moderating variables (e.g., perceptions of low competence, high fear of failure, achievement contexts with high evaluative criteria) between performance-approach and performance-avoidance goal adoption. Perceived ability and perceived difficulties when engaged in an achievement task have been discussed by several authors. They were mentioned by Nicholls (1984) and Dweck (1986) as potential factors involved in the transformation of an originally adopted performance-approach to a performance-avoidance orientation. More recently, Kaplan, *et al.* (2002), commenting on relevant research results of their own, remarked that the shifting of a performance-approach to a performance-avoidance goal orientation may be connected to changing circumstances related to the achievement context or to experienced difficulties. These circumstances may serve as a cue signaling that the valued exhibition of high ability is in danger and directing the person towards the adoption of a performance-avoidance goal.

The weak correlation between mastery-avoidance and performance-approach goals was theoretically anticipated, because there is no conceptual overlap between these goals neither on the definition or the valence dimension of competence.

One important concern was related to the sample size required for conducting CFA research. In the case of confirmatory factor analysis, the understanding and the interpretation of the research results may be compromised by the sample size. But, according to the calculations tabulated by Gagné and Hancock (2006), the estimates of the four-factor structure model were reasonably accurate. This derives from the consideration of

the excellent reliability estimates, along with the estimated standardized factor loadings. However, it would be interesting and useful for future research in Greece and elsewhere to further examine, in larger samples with more equally balanced gender representation, the psychometric properties of the different achievement goal measures considered in the present research attempt. Such an examination would help in clarifying the factor structure of achievement goals.

REFERENCES

- ABD-EL-FATTAH, S. M., & AL-NABHANI, H. Z. (2012) From self-theories of intelligence to academic delay of gratification: the mediating role of achievement goals. *Australian Journal of Educational & Developmental Psychology*, 12, 93-107.
- ALKHARUSI, H., & ALDHAFRI, S. (2010) Gender differences in the factor structure of the 2 × 2 achievement goal framework. *College Student Journal*. Retrieved from <http://www.faqs.org/periodicals/201009/2151752371.html>.
- AMES, C. (1992) Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271.
- AMES, C., & ARCHER, J. (1988) Achievement goals in the classroom: students' learning strategies and motivational processes. *Journal of Educational Psychology*, 80, 260-267.
- BAGOZZI, R. P., & YI, Y. (1988) On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74-94.
- BARKOUKIS, V., NTOUMANIS, N., & NIKITARAS, N. (2007) Comparing dichotomous and trichotomous approaches to achievement goal theory: an example using motivational regulations as outcome variables. *British Journal of Educational Psychology*, 77, 683-702.
- BARKOUKIS, V., THØGERSEN-NTOUMANI, C., NTOUMANIS, N., & NIKITARAS, N. (2007) Achievement goals in physical education: examining the predictive ability of five different dimensions of motivational climate. *European Physical Education Review*, 13, 267-285.
- BARRON, K. E., BARANICK, L. E., & FINNEY, S. J. (2006) Mastery avoidance at work. Paper presented at the American Educational Research Association's National Conference, San Francisco, CA, April 7-11.
- BENTLER, P. M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- BOEKAERTS, M. (2001) Context sensitivity: activated motivational beliefs, current concerns and emotional arousal. In S. Volet & S. Järvelä (Eds.), *Motivation in learning contexts: theoretical advances and methodological implications*. Oxford, UK: Elsevier Science Ltd. Pp. 17-31.
- BROWNE, M. W., & CUDECK, R. (1993) Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Beverly Hills, CA: Sage. Pp. 136-162.
- CAMPBELL, H. L., BARRY, C. L., & FINNEY, S. J. (2008) Configural, metric, and scalar invariance of the modified achievement goal questionnaire across African American and White university students. *Educational and Psychological Measurement*, 68, 988-1007.

- CARETTE, B., ANSEEL, F., & VAN YPEREN, N. W. (2011) Born to learn or born to win? Birth order effects on achievement goals. *Journal of Research in Personality*, 45, 500-503.
- CONROY, D. E., & ELLIOT, A. J. (2004) Fear of failure and achievement goals in sport: addressing the issue of the chicken and the egg. *Anxiety, Stress, and Coping*, 17, 271-285.
- DAVID, D. M. M. (2009) Achievement goals in mathematics. *The Assessment Handbook*, 2, 53-63.
- DERMITZAKI, I., & EFKLIDES, A. (2001) Age and gender effects on students' evaluations regarding the self and task-related experiences in mathematics. In S. Volet & S. Järvelä (Eds.), *Motivation in learning contexts: theoretical advances and methodological implications*. Oxford, UK: Elsevier. Pp. 271-293.
- DESHON, R. P., & GILLEPSIE, J. Z. (2005) A motivated action theory account of goal orientation. *Journal of Applied Psychology*, 90, 1096-1127.
- DUPEYRAT, C., & MARINÉ, C. (2005) Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: a test of Dweck's model with returning to school adults. *Contemporary Educational Psychology*, 30, 43-59.
- DWECK, C. S. (1986) Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.
- DWECK, C. S., & BEMPECHAT, J. (1983) Children's theories of intelligence: consequences for learning. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), *Learning and motivation in the classroom*. Hillsdale, NJ: Lawrence Erlbaum. Pp. 239-256.
- DWECK, C. S., & LEGGETT, E. L. (1988) A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- DWECK, C. S., & MOLDEN, D. C. (2005) Self-theories: their impact on competence motivation and acquisition. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford Press. Pp. 122-140.
- DYKMAN, B. M. (1998) Integrating cognitive and motivational factors in depression: initial tests of a goal orientation approach. *Journal of Personality and Social Psychology*, 74, 139-158.
- ELLIOT, A. J. (1997a) A hierarchical model of approach and avoidance motivation. *Journal of Personality and Social Psychology*, 72, 218-232.
- ELLIOT, A. J. (1997b) Integrating "classic" and "contemporary" approaches to achievement motivation: a hierarchical model of approach and avoidance achievement motivation. In P. Pintrich & M. Maehr (Eds.), *Advances in motivation and achievement*. Vol. 10. Greenwich, CT: JAI Press. Pp. 143-179.
- ELLIOT, A. J. (1999) Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34, 169-189.
- ELLIOT, A. J. (2005) A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford Press. Pp. 52-72.
- ELLIOT, A. J. (2006) The hierarchical model of approach-avoidance motivation. *Motivation and Emotion*, 30, 111-116.
- ELLIOT, A. J. (2008) Goals. In N. J. Salkind (Ed.), *Encyclopedia of educational psychology*. Vol. 1. Thousand Oaks, CA: Sage. Pp. 443-445.
- ELLIOT, A. J., & CHURCH, M. A. (1997) A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 77, 218-232.

- ELLIOT, A. J., & COVINGTON, M. V. (2001) Approach and avoidance motivation. *Educational Psychology Review*, 13, 73-91.
- ELLIOT, A. J., & HARACKIEWICZ, J. M. (1996) Approach and avoidance achievement goals and intrinsic motivation: a meditational analysis. *Journal of Personality and Social Psychology*, 70, 461-475.
- ELLIOT, A. J., & MCGREGOR, H. A. (2001) A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-519.
- ELLIOT, A. J., MCGREGOR, H. A., & THRASH, T. M. (2002) The need for competence. In E. Deci & R. Ryan (Eds.), *Handbook of self-determination theory*. Rochester, NY: University of Rochester Press. Pp. 361-387.
- ELLIOT, A. J., & MURAYAMA, K. (2008) On the measurement of achievement goals: critique, illustration, and application. *Journal of Educational Psychology*, 100, 613-628.
- ELLIOT, A. J., & REIS, H. (2003) Attachment and exploration in adulthood. *Journal of Personality and Social Psychology*, 85, 317-331.
- ELLIOT, A. J., & THRASH, T. M. (2001) Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review*, 13, 139-156.
- ELLIOTT, E. S., & DWECK, C. S. (1988) Goals: an approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5-12.
- ENDERS, C. K., & TOFIGHI, D. (2008) The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15, 75-95.
- FINNEY, S. J., PIEPER, S. L., & BARRON, K. E. (2004) Examining the psychometric properties of the Achievement Goal Questionnaire in a general academic context. *Educational and Psychological Measurement*, 64, 365-382.
- GAGNÉ, P., & HANCOCK, G. R. (2006) Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, 41, 65-83.
- GRAHAM, S., & HUDLEY, C. (2005) Race and ethnicity in the study of motivation and competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford. Pp. 392-413.
- GRANT, H., & DWECK, C. S. (2003) Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85, 541-553.
- HAMBLETON, R. K., & KANJEE, A. (1995) Increasing the validity of cross-cultural assessments: use of improved methods for test adaptations. *European Journal of Psychological Assessment*, 11, 147-157.
- HARACKIEWICZ, J. M., BARRON, K. E., & ELLIOT, A. (1998) Rethinking achievement goals: when are they adaptive for college students and why? *Educational Psychologist*, 33, 1-21.
- HARACKIEWICZ, J. M., & LINNENBRINK, E. A. (2005) Multiple achievement goals and multiple pathways for learning: the agenda and impact of Paul R. Pintrich. *Educational Psychologist*, 40, 75-84.
- HOYLE, R., & PANTER, A. (1995) Writing about structural equation models. In R. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications*. Thousand Oaks, CA: Sage. Pp. 100-119.
- HU, L., & BENTLER, P. M. (1995) Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues and applications*. Thousand Oaks, CA: Sage. Pp. 76-99.

- HU, L., & BENTLER, P. M. (1998) Fit indices in covariance structure analysis: sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- INGLES, C. J., GARCIA-FERNANDEZ, J. M., CASTEJON, J. L., VALLE, A., DELGADO, B., & MARZO, J. C. (2009) Reliability and validity evidence of scores on the achievement goal tendencies questionnaire in a sample of Spanish students of compulsory secondary education. *Psychology in the Schools*, 46, 1048-1060.
- JÖRESKOG, K. G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- KAPLAN, A., & MAEHR, M. L. (2007) The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19, 141-184.
- KAPLAN, A., MIDDLETON, M. J., URDAN, T., & MIDGLEY, C. (2002) Achievement goals and goal structures. In C. Midgley (Ed.), *Goals, goal structures, and patterns of adaptive learning*. Mahwah, NJ: Erlbaum. Pp. 21-53.
- KOULI, O., & PAPAIOANNOU, A. G. (2009) Ethnic/cultural identity salience, achievement goals and motivational climate in multicultural physical education classes. *Psychology of Sport and Exercise*, 10, 45-51.
- KUNDA, Z. (1990) The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- LAU, K. L., & LEE, J. C. K. (2008) Validation of a Chinese achievement goal orientation questionnaire. *British Journal of Educational Psychology*, 78, 331-353.
- LEONTARI, A., & GIALAMAS, V. (2002) Implicit theories, goal orientations and perceived competence: impact on students' achievement behaviour. *Psychology in the Schools*, 39, 279-291.
- LEWIN, K. (1935) *A Dynamic Theory of Personality*. New York: McGraw-Hill.
- MAEHR, M. L., & MIDGLEY, C. (1991) Enhancing student motivation: a school-wide approach. *Educational Psychologist*, 26, 399-427.
- MCINERNEY, D. M., ROCHE, L. A., MCINERNEY, V., & MARSH, H. W. (1997) Cultural perspectives on school motivation: the relevance and application of goal theory. *American Educational Research Journal*, 34, 207-236.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. I. (1953) *The achievement motive*. New York: Appleton-Century-Crofts.
- MESSICK, S. (1989) Validity. In R. L. Linn (Ed.), *Educational measurement*. (3rd ed.) New York: American Council on Education/Macmillan. Pp. 13-103.
- MIDGLEY, C., KAPLAN, A., & MIDDLETON, M. (2001) Performance-approach goals: good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93, 77-86.
- MURAYAMA, K., ZHOU, M., & NESBIT, J. C. (2009) A cross-cultural examination of the psychometric properties of responses to the achievement goal questionnaire. *Educational and Psychological Measurement*, 69, 266-286.
- MUTHÉN, B. (1989a) Factor structure in groups selected on observed scores. *British Journal of Mathematical and Statistical Psychology*, 42, 81-90.
- MUTHÉN, B. (1989b) Multiple-group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, 42, 55-62.
- NICHOLLS, J. G. (1984) Achievement motivation: conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.
- OMMUNDSEN, Y. (2001) Self-handicapping strategies in physical education classes: the influence of implicit theories of the nature of ability and achievement goal orientations. *Psychology of Sport and Exercise*, 2, 139-156.

- PAYNE, S. C., YOUNGCOURT, S. S., & BEAUBIEN, J. M. (2007) A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92, 128-150.
- PINTRICH, P. R. (2000) An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25, 92-104.
- ROESER, R. W. (2004) Competing schools of thought in achievement goal theory? In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement*. Vol. 13. *Motivating students, improving schools: the legacy of Carol Midgley*. Oxford, UK: Elsevier. Pp. 265-300.
- SATORRA, A. (2000) Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: a Festschrift for Heinz Neudecker*. London, UK: Kluwer Academic Publishers. Pp. 233-247.
- SIDERIDIS, G. D. (2005) Goal orientation, academic achievement, and depression: evidence in favour of a revised goal theory framework. *Journal of Educational Psychology*, 97, 366-375.
- SIDERIDIS, G. D. (2006a) Achievement goal orientations, "oughts," and self-regulation in students with and without learning disabilities. *Learning Disability Quarterly*, 29, 3-18.
- SIDERIDIS, G. D. (2006b) Goal orientations and strong oughts: adaptive or maladaptive forms of motivation for students with and without suspected learning disabilities. *Learning and Individual Differences*, 16, 61-77.
- STEVENSON, S., & LOCHBAUM, M. (2008) Understanding exercise motivation: examining the revised social-cognitive model of achievement motivation. *Journal of Sport Behaviour*, 31, 389-412.
- SCLOVE, S. L. (1987) Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- TOFIGHI, D., & ENDERS, C. K. (2007) Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*. Greenwich, CT: Information Age. Pp. 317-341.
- URDAN, T. (1997) Achievement goal theory: past results, future directions. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement*. Vol. 10. Greenwich, CT: JAI Press. Pp. 99-141.
- URDAN, T. (2004) Predictors of academic self-handicapping and achievement: examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology*, 96, 251-264.
- VANDEWALLE, D. (1997) Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement*, 57, 995-1015.
- VAN YPEREN, N. (2003) Task interest and actual performance: the moderating effects of assigned and adopted purpose goals. *Journal of Personality and Social Psychology*, 85, 1006-1015.
- VAN YPEREN, N. (2006) A novel approach to assessing achievement goals in the context of the 2×2 framework: identifying distinct profiles of individuals with different dominant achievement goals. *Personality and Social Psychology Bulletin*, 32, 1432-1445.
- VOLET, S. (2001) Emerging trends in recent research on motivation in learning contexts. In S. Volet & S. Järvelä (Eds.), *Motivation in learning contexts: theoretical advances and methodological implications*. Oxford, UK: Elsevier Science Ltd. Pp. 319-334.

- VOLET, S., & JÄRVELÄ, S. (Eds.). (2001) *Motivation in learning contexts: theoretical advances and methodological implications*. Oxford, UK: Elsevier Science Ltd.
- WANG, C. K. J., BIDDLE, S. J. H., & ELLIOT, A. J. (2007) The 2×2 achievement goal framework in a physical education context. *Psychology of Sport and Exercise*, 8, 147-168.
- YU, S. L., & WOLTERS, C. A. (2002) Issues in the assessment of motivation in students from ethnic minority populations. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement*. Vol. 12. Oxford, UK: Elsevier Science Ltd. Pp. 349-380.

Accepted September 13, 2013.

APPENDIX

Greek Achievement Goal Questionnaire-Revised (AGQ-R) Items (Elliot & Murayama, 2008).

1. Ο σκοπός μου είναι να μάθω τέλεια την ύλη που παρουσιάζεται σε αυτό το μάθημα.
2. Πασχίζω να κατανοήσω το περιεχόμενο αυτού του μαθήματος όσο το δυνατόν πληρέστερα.
3. Ο στόχος μου είναι να μάθω όσο το δυνατόν περισσότερα.
4. Ο σκοπός μου είναι να αποφύγω να μάθω λιγότερα από όσα πιθανόν θα μπορούσα.
5. Πασχίζω να αποφύγω μια ελλιπή κατανόηση της ύλης αυτού του μαθήματος.
6. Ο στόχος μου είναι να αποφύγω να μάθω λιγότερα από όσα είναι δυνατό να μάθω.
7. Ο σκοπός μου είναι να τα πάω καλά σε σχέση με άλλες-ους φοιτήτριες-ές.
8. Πασχίζω να τα πάω καλά σε σύγκριση με άλλες-ους φοιτήτριες-ές.
9. Ο στόχος μου είναι να τα πάω καλύτερα απ'ό,τι οι άλλες-οι φοιτήτριες-ές.
10. Ο σκοπός μου είναι να αποφύγω να τα πάω χειρότερα απ'ό,τι άλλες-οι φοιτήτριες-ές.
11. Πασχίζω να αποφύγω να τα πάω χειρότερα απ'ό,τι άλλες-οι.
12. Ο στόχος μου είναι να αποφύγω να τα πάω άσχημα σε σύγκριση με άλλες-ους.

BRAND EMOTIONAL CREDIBILITY: EFFECTS OF MIXED EMOTIONS ABOUT BRANDED PRODUCTS WITH VARYING CREDIBILITY^{1, 2}

ANTONIO MILETI, M. IRENE PRETE, AND GIANLUIGI GUIDO

*Department of Management, Economics, Mathematics and Statistics
University of Salento, Lecce*

Summary.—This research investigates the effects of mixed emotions on the positioning and on the intention to purchase different categories of branded products (i.e., Attractiveness-products, Expertise-products, and Trustworthiness-products), in relation to their main component of credibility (Ohanian, 1990). On the basis of a focus group ($n = 12$) aimed to identify the three branded products used as stimuli and a pre-test ($n = 240$) directed to discover emotions elicited by them, two studies ($n = 630$; $n = 240$) were carried out. Positioning and multiple regression analyses showed that positive and negative emotions are positively related with the positioning and the purchase intention of Attractiveness-products, and, respectively, positively and negatively related with those of Trustworthiness-products; whereas negative emotions are negatively associated with those of Expertise-products. Brand Emotional Credibility—i.e., the emotional believability of the brand positioning signals—may help to identify unconscious elements and the simultaneous importance of mixed emotions associated with different products to match consumers' desires and expectations.

Consumer choices are often motivated by intense, mixed, and opposing emotional states (Hirschman & Stern, 1999; Williams & Aaker, 2002): individuals may feel guilty about buying an expensive good, ashamed of being associated with an embarrassing product, or feel pleased about flaunting a fashion brand. The interest of marketing literature in consumers' emotional experiences has two main motivations: on the one hand, the attention to examine subconscious influences on consumer choices (Bagozzi, Gopinath & Nyer, 1999); and, on the other hand, the recent interest in studying the coexistence of both positive and negative emotions (Otnes, Lowrey, & Shrum, 1997; Ruth, Brunel, & Otnes, 2002; Williams & Aaker, 2002; Penz & Hogg, 2011; Aviezer, Trope, & Todorov, 2012). Emotions, defined as states of readiness to act which arise from the evaluation of events or thoughts, together with physiological processes and behavioral manifestations (Bagozzi, *et al.*, 1999), are crucial in purchase settings. Therefore, consumption emotions can be delineated as feeling states experienced when considering, purchasing, or consuming products (Cohen & Areni, 1991).

¹Address correspondence to Prof. Gianluigi Guido, Department of Management, Economics, Mathematics and Statistics, University of Salento, Via per Monteroni, Palazzo Ecotekne, 73100 Lecce, Italy or e-mail (gianluigi.guido@unisalento.it).

²The authors would like to thank Nunzia Marra and Tanja Scarlino for their kind support in data collection and codification.

Opposite emotions, positive and negative ones, coexist at various levels. Neurobiological literature shows shared mechanisms to control conflicting emotions, such as pain and pleasure (Leknes & Tracey, 2008). Psychologists suggest that opposing emotions can concurrently occur in everyday circumstances (Cacioppo, Gardner, & Berntson, 1997; Larsen, McGraw, & Cacioppo, 2001). In consumption settings, opposite emotions, besides influencing consumers' attitudes (Ruth, 2001; Williams & Aaker, 2002) and product preferences (Chitturi, Raghunathan, & Mahajan, 2007), have a critical importance for technological products (Mick & Fournier, 1998) and impulsive buying experiences (Rook, 1985; Ramanathan & Williams, 2007).

Emotions, deemed to be a relevant brand attribute and a signal of product positioning (Haley, 1985; Ruth, 2001), have been of interest to other areas such as brand credibility (Maathuis, Rodenberg, & Sikkels, 2004), which is defined as the "believability of the product position information embedded in a brand" (Erdem & Swait, 2004, p. 191). Brand credibility encloses the following dimensions: Attractiveness, the degree of (not only external) pleasantness assigned; Expertise, the degree of experience gained in a specific field of knowledge; and Trustworthiness, the degree of given trust (Hovland, Janis, & Kelley, 1953; McGuire, 1983; Ohanian, 1990). Developed to deal with celebrities endorsers, Ohanian's (1990) scale contemplates these dimensions, and it is considered a well-known and reliable measure of product credibility—e.g., for mobile advertising (Prete, 2007), or branded products (Wang & Yang, 2010, 2011; Li, Wang, & Yang, 2011).

This research represents an advance of existing studies in brand credibility and intends to investigate the effects of mixed emotions—both positive and negative—on the positioning and the intention to purchase different categories of branded products, regarded to have varying credibility. "Branded products" were used so as not to separate the brand from its product category, thus assessing the joint impact on perception (e.g., the brand "Rolex" was not assessed separately from the category of "watches," but "Rolex watches," "Levi's jeans," or "Volvo cars," and so on). "Varying credibility" identifies the different importance of each of the aforementioned components of credibility: Attractiveness, Expertise, or Trustworthiness.

Research by Maathuis, *et al.* (2004) has highlighted the relevance of cognitive reasons for the credibility of grocery branded products (e.g., detergents and food) and, conversely, the importance of emotional reasons for the credibility of shopping (e.g., clothing and fashion) and high technology (e.g., telecommunication companies and hi-fi) branded products. Furthermore, it has been showed that Attractiveness is relevant for those categories of products whose image is influenced by external aspects, and

is linked to style and fashion. Expertise is significant for products whose choice is focused on their adequacy and effectiveness in solving specific problems; as for example, products for personal hygiene and pharmaceuticals. Finally, Trustworthiness has an impact on products whose image is mainly linked to the concept of resistance and durability, as is the case for technology products (Guido & Peluso, 2006, 2009, 2011).

This study's research has a twofold aim. The first purpose concerns the evaluation of the different mix of emotions that branded products evoke in relation to their main component of credibility (i.e., Attractiveness, Expertise, and Trustworthiness). A positioning analysis—a well-known method of investigating how a brand is perceived by the target market (Wind, 1990)—is carried out, by using emotions as dimensions that define the product 'perceptual space' in the consumers' mind. The second purpose concerns the investigation of the relationship among emotions (positive and negative) and the intention to purchase the different categories of products (i.e., Attractiveness-products, Expertise-products, and Trustworthiness-products).

Study 1

POSITIONING OF BRANDS WITH VARYING CREDIBILITY

The purpose of Study 1 is to evaluate emotions evoked by different categories of branded products in relation to their main component of credibility, by placing these branded products on the consumers' emotional perceptual space.

Study 1 is based on the assumption that the value of emotions elicited, positive or negative, has different implications, depending on the product category considered. Both positive and negative emotions can influence the evaluation of products with a hedonistic nature such as Attractiveness-products, which are characterized by multi-sensory aspects, where the consumer is conscious of the emotive shopping experience (e.g., clothing, cosmetics, and perfumes). On the contrary, both positive and negative emotions can have a lower value for products with a mainly utilitarian nature, which are designed to satisfy predominantly functional, logical, or instrumental needs such as Expertise-products (e.g., products for hygiene, tools, and utensils). Finally, there is a mixed type of link between emotions and products that must last over time, such as Trustworthiness-products (e.g., high-tech products or electrical appliances) because, while it is possible that these products, with their characteristics of innovation, will generally evoke positive emotions in potential buyers, it is more difficult for them to evoke negative emotions since their assessment is usually based on the properties of efficiency and effectiveness (Maathuis, *et al.*, 2004; Guido & Peluso, 2006, 2009, 2011).

Therefore, the following hypotheses were proposed:

Hypothesis 1a. Attractiveness-branded products should be positioned in a segment of the consumers' perceptual map characterized by a high level of both positive and negative emotions.

Hypothesis 1b. Expertise-branded products should be positioned in a segment of the consumers' perceptual map characterized by a low level of both positive and negative emotions.

Hypothesis 1c. Trustworthiness-branded products should be positioned in a segment of the consumers' perceptual map characterized by a high level of positive emotions rather than negative ones.

Pilot Study

STIMULUS AND QUESTIONNAIRE DEVELOPMENT

Sample

To identify the three types of products and their corresponding brands able to represent a particular component of credibility (Ohanian, 1990), a focus group ($n = 12$; 50% men; M age = 22.5 yr.; 50% graduate students, 50% undergraduates) was carried out. Results of the content analysis showed the selection of the following categories of products: underwear, as a product representative of Attractiveness; toothpaste, as a product representative of Expertise; and a cell phone, as a product representative of Trustworthiness. In the same way, the brands Intimissimi, Mentadent, and Nokia, three well-known brands in the Italian market, were respectively identified. A pretest on a sample of 240 individuals was carried out (80 per product) (39.1% men; M age = 23.5 yr., all unmarried and without children).

Questionnaire Development

Two apparently conflicting criteria guided the selection of a representative and comprehensive set of emotions (cf., Bagozzi, *et al.*, 1999; Laros & Steenkamp, 2005). On the one hand, there was the need to represent the full range of emotive experiences that are relevant for the branded products in question and not just for one specific area. On the other hand, it was necessary to focus on a small number of emotions in order to use a measure that is both efficient and economical. The following taxonomies were selected: Izard's (1977) Differential Emotions Theory, Plutchik's (1980) Circular Model of Emotion—which consider basic or primary emotions as innate and universal—and a more recent taxonomy put forward by O'Shaughnessy & O'Shaughnessy (1999), an adaptation of Gordon's (1987) "Belief-Wish-Condition" model (BWC Model), which proposes a

more detailed structure of the emotions in reference to marketing. This approach allows consideration of not only the primary emotions felt by all humans, but also the secondary or social emotions linked to the sociocultural environment (cf. Damasio, 1999).

Specifying the time and place for the purchase (cf. Fishbein & Ajzen, 1975), participants were asked to list the emotional aspects linked to their purchase of the three branded products relating to: Anger, Disgust, Contempt, Joy, Interest, Fear, Guilt, Surprise, Sadness, and Shame, for Izard's (1977) taxonomy; Acceptance, Expectancy, Anger, Disgust, Joy, Fear, Surprise, and Sadness for Plutchik's (1980) taxonomy; Love, Anger/Outrage, Embarrassment, Integrity, Pride, Fear, Guilt, Hope, and Shame, for O'Shaughnessy and O'Shaughnessy's (1999) taxonomy. For example, respondents were asked: "Imagine that in three to six months you are going to buy some Intimissimi underwear at your favorite shop, what might your fears be?" or "What could be the aspects that could make you feel hopeful, if you were to buy in three-six months some Mentadent toothpaste at your favorite store?" or "What aspects might make you feel surprised if in three to six months you bought a Nokia mobile phone at your favorite shop?"

Results

Respondents were recruited by students in two Italian major universities as an extra credit assignment, and the response rate was 80%. An open-ended questionnaire was administered in order to elicit the emotions relating to the purchase of the three considered branded products. A content analysis was performed by considering the occurrence of high frequency answers with a minimum of eight citations per emotion, which were used in the preparation of three main close-ended questionnaires, one for each branded products.

Study 1

METHOD

Sample

The main study was conducted using a convenience sample of 630 people (210 per product; 34.3% men; *M* age = 23.5 yr.; all unmarried and without children).

Materials and Procedure

The questionnaire contained a screening question to confirm the knowledge of each brand, to make sure that only those interviewees knowing the considered brand participated in the survey. According to the settings of the expectancy-value models (cf., Fishbein & Ajzen, 1975), the probability the described event could happen and the importance of the expected consequences, were measured on a 7-point Likert-type scale,

with anchors 1: Not at all and 7: Very much, by asking "Could you please evaluate the probability of the following facts happening and the importance of the expected consequences?" (e.g., "The toothpaste is not efficacious," "The toothpaste has a good taste") Although various methods could be used, self-evaluation of one's own emotive experiences is one of the most common in academic research (Bagozzi, *et al.*, 1999). Socio-demographic questions were also included.

Participants were recruited by students in a major Italian university as an extra credit assignment, and were administered the closed-end questionnaires. They were assured their answers were anonymous, they could withdraw at any time, there were no right or wrong responses, and that they should answer questions as honestly as possible.

Analyses

Reliability of the questionnaires was evaluated by measuring internal consistency using Cronbach's α (Attractiveness-products: $\alpha = .86$; Expertise-products: $\alpha = .89$; Trustworthiness-products: $\alpha = .86$) and correlations among items ($> .30$). An exploratory factor analysis, including the 18 considered emotions, was carried out ($N = 630$; subject-to-variable ratio = 35:1) using maximum likelihood extraction followed by Varimax rotation and assessment of the Scree test and coherence of item meanings within factors (Table 1). The two factors identified, Positive Emotions and Negative Emotions, were chosen as positioning variables and allowed to place the three branded products on a bi-dimensional perceptual map, which is a marketing tool used to visually display a brand's position in the consumers' minds.

RESULTS

The results support Hypotheses 1a, 1b, and 1c, since the Attractiveness-product (Intimissimi) was positioned in a segment with high emotivity (positive and negative), the Expertise-product (Mentadent) in a segment with low emotivity (positive and negative), and the Trustworthiness-product (Nokia) in a segment characterized by high positive emotivity and low negative emotivity (Table 1, Fig. 1).

Study 2

RELATIONS BETWEEN OPPOSING EMOTIONS AND INTENTION TO BUY BRANDED PRODUCTS WITH VARYING CREDIBILITY

Consumers express their purchasing emotive experiences on the basis of positive and/or negative emotions, according to the classic bi-polarization of this construct (Laros & Steenkamp, 2005). It was posited that the relationship between these emotions and the intention to buy specific branded products would differ according to their prevalent dimension of credibility. Attractiveness products allow consumers to obtain promo-

TABLE 1
FACTOR LOADING OF EMOTION RATINGS FROM THREE TAXONOMIES (IZARD; PLUTCHIK;
O'SHAUGHNESSY AND O'SHAUGHNESSY)

	<i>M</i>	<i>SD</i>	Factor Loading	Communality
Positive Emotions				
Acceptance	144.52	49.34	.84	0.72
Interest	138.32	51.94	.83	0.70
Expectancy	154.77	51.12	.80	0.65
Joy	136.68	51.41	.78	0.61
Integrity	147.94	48.35	.76	0.57
Love	154.40	42.21	.74	0.55
Surprise	106.80	39.76	.73	0.56
Pride	159.60	55.81	.70	0.50
Hope	178.04	54.85	.44	0.20
Negative Emotions				
Contempt	104.14	43.69	.77	0.60
Disgust	92.47	40.59	.76	0.59
Sadness	112.05	49.58	.73	0.58
Remorse	88.85	38.55	.66	0.44
Anger	113.88	51.28	.63	0.41
Guilt	91.68	37.01	.61	0.38
Fear	77.48	30.85	.52	0.27
Shame	76.63	41.00	.51	0.27
Embarrassment	85.38	41.55	.50	0.25

Note.—Eigenvalues of the factors after rotation: Positive Emotions = 29.60; Negative Emotions = 19.45; Total amount of variance explained = 49.06%.

tion goals to achieve pleasant consumption experience through, e.g., Joy related to “looking fashionable,” or Fear associated to the “non-availability of the favorite item in one’s own size,” thus suggesting the leading role of emotional aspects of consumption. Expertise products are more related to prevention goals—to avoid distressful consumption experience through, e.g., Integrity related to “being conscientious,” or through, e.g., Fear related to the “low efficacy of the product”—thus evoking the importance of rational and utilitarian aspects of consumption (Higgins, 1997). Trustworthiness products allow a person to achieve both promotion and prevention goals, often producing typical paradoxes of technological products. For example, these goods can facilitate or obstruct order, provide benefits of scientific knowledge or produce obsolescence (Mick & Fournier, 1998), thus evoking opposing emotions such as, e.g., Joy connected to buy a “high quality mobile phone,” or Fear related to an “insuffi-

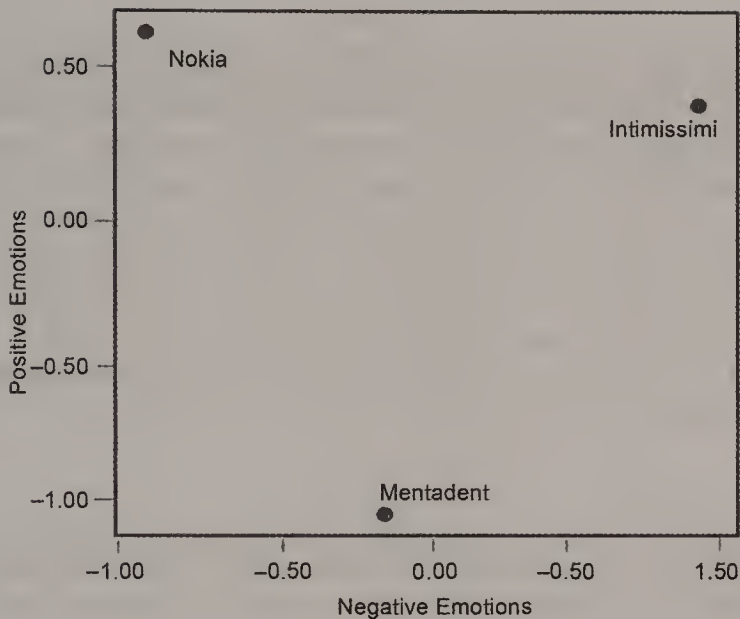


FIG. 1. Positioning of the three brands according to the joint emotions of the three taxonomies (Izard; Plutchik; O'Shaughnessy and O'Shaughnessy)

cient price/quality ratio." It can be assumed that the relationship with the intention to purchase is similarly characterized, therefore:

Hypothesis 2a. Both positive and negative emotions should account for significant variance in the intention to purchase Attractiveness-branded products, producing a positive relationship.

Hypothesis 2b. Negative, but not positive, emotions should account for significant variance in the intention to purchase Expertise-branded products, producing a negative relationship.

Hypothesis 2c. Both positive and negative emotions should account for significant variance in the intention to purchase Trustworthiness-branded products, producing a mixed relationship.

METHOD

A replication-extension study was carried out, combining results from Study 1 with results of this new study particularly intended to replicate and extend previous findings.

Sample

A convenience sample was used, with 240 students from a major university in Italy, 80 per product (39.2% men; M age = 25 yr.; all unmarried and without children). Participants were recruited by students in a Marketing course at a major Italian university as an extra credit assignment.

Measures

The questionnaire used repeated the questions in Study 1, and was integrated with two additional questions designed to measure the "Intention to purchase," as the result of the product between the strength of the intention and the probability of purchasing a specific product at a preferred point-of-sale over the next 3 to 6 months. As in Study 1, measures were created according to the expectancy-value model (Fishbein & Ajzen, 1975). For example, participants were asked, on a 7-point Likert-type scale: "If you want to purchase a mobile phone, how likely is it that you buy a Nokia?" and "If you want to purchase a mobile phone, how strong is your intention to buy a Nokia?"

Analyses

For each of the considered branded products, an exploratory factor analysis ($N=80$; subject-to-variable ratio=4.4:1; maximum likelihood extraction) was run. Criteria for the determination of factors were the Scree test and the coherence of item meanings within factors. The rotation method was Oblimin. The analysis jointly considered the emotions related to the three taxonomies. The two latent factors obtained were used as independent variables in three regression analyses, one for each product, to study their relationships with the dependent variable "Intention to purchase."

RESULTS

With regard to the Attractiveness-product Intimissimi, the factor analysis identified two latent factors, i.e., Positive Emotions and Negative Emotions (Table 2). Positive Emotions ($\beta = 0.40, p < .01$) and Negative Emotions ($\beta = 0.50, p < .01$) account for significant variance in the Intention to purchase the brand (Table 3). Not only positive emotions can lead to an incentive to buy products, but also negative emotions such as fear—for example, not being up to date with current times or with fashion; or shame—such as feeling excluded from a certain social elite or from a reference group or one the consumer is part of. Hypothesis 2a was supported.

For the Expertise-product Mentadent, the factor analysis resulted also in the identification of two latent factors, i.e., Negative Emotions and Positive Emotions (Table 4). Only Negative Emotions ($\beta = -0.47, p < .01$), but not Positive ones ($p < .05$), were associated with the intention to purchase (Hypothesis 2b), and Negative Emotions actually discourage purchase (Table 5). This means the rational aspects that dominate the evaluation of Expertise-products take precedence over emotional ones, when deciding to purchase. Thus, only negative emotions are related to the purchase and these could inhibit buying something to satisfy a utilitarian and rational rather than a recreational or leisure need.

TABLE 2
FACTOR LOADINGS OF EMOTION RATINGS FROM THREE TAXONOMIES (IZARD; PLUTCHIK;
O'SHAUGHNESSY AND O'SHAUGHNESSY) CONCERNING THE ATTRACTIVENESS-PRODUCT

Positive Emotion	Factor Loading	Communalities	Negative Emotion	Factor Loading	Communalities
Love	.90	0.76	Sadness	.89	0.83
Joy	.89	0.76	Contempt	.87	0.75
Expectancy	.85	0.74	Remorse	.84	0.65
Acceptance	.84	0.68	Disgust	.76	0.56
Interest	.82	0.65	Guilt	.68	0.45
Integrity	.81	0.75	Anger	.61	0.43
Surprise	.72	0.69	Embarrassment	.41	0.16
Hope	.67	0.42	Shame	.37	0.16
Pride	.64	0.55	Fear	.14	0.10

Note.—Eigenvalues of the factors after rotation: Positive emotions = 38.85; Negative emotions = 17.54; Total amount of variance explained = 55.59%.

Finally, for the Trustworthiness-product Nokia, factor analysis reproduced the results of Study 1, identifying the two underlying factors of emotions (Table 6). The relationship between the emotive variables and the intention to purchase—Positive Emotions ($\beta = 0.61$, $p < .001$) and Negative Emotions ($\beta = -0.59$, $p < .001$)—on the one hand is consistent with Hypothesis 2c, and on the other reinforces the findings related to the previous hypothesis. As for Expertise-products, Negative Emotions scores had a negative relationship with the intention to conclude the purchase of these categories of goods targeted to solve specific problems (Table 7).

TABLE 3
INFLUENCE OF NEGATIVE AND POSITIVE EMOTION RATINGS FROM THREE TAXONOMIES (IZARD;
PLUTCHIK; O'SHAUGHNESSY AND O'SHAUGHNESSY) FOR THE ATTRACTIVENESS-PRODUCT

Variable	B	Std. Error	Beta	t
Constant	19.58	0.79		24.91†
Positive Emotion	4.14	0.86	0.40	4.82†
Negative Emotion	5.26	0.87	0.48	6.05†

Note.— $N = 80$. The dependent variable was Intention to purchase ($R = .73$; $R^2 = .54$; $\text{Adj. } R^2 = .52$; standard error = 7.03; $F_{2,79} = 44.47$, $p < .001$). † $p < .01$.

GENERAL DISCUSSION

This research may represent a step forward in the study of consumer behavior. An emotional mapping of different categories of products, on the basis of their main element of credibility, was created and the different

TABLE 4
FACTOR LOADINGS OF EMOTION RATINGS FROM THREE TAXONOMIES (IZARD; PLUTCHIK;
O'SHAUGHNESSY AND O'SHAUGHNESSY) CONCERNING THE EXPERTISE-PRODUCT

Negative Emotion	Factor Loading	Commun- alities	Positive Emotion	Factor Loading	Commun- alities
Anger	.79	0.61	Joy	.80	0.76
Shame	.77	0.62	Interest	.77	0.68
Embarrassment	.76	0.60	Acceptance	.76	0.58
Sadness	.72	0.54	Expectancy	.59	0.35
Guilt	.72	0.52	Surprise	.56	0.46
Remorse	.69	0.49	Hope	.49	0.25
Contempt	.66	0.46	Pride	.48	0.23
Disgust	.51	0.26	Love	.36	0.13
Fear	.27	0.08	Integrity	0.30	0.11

Note.—Eigenvalues of the factors after rotation: Negative Emotions = 25.93; Positive Emotions = 16.99; Total amount of variance explained = 42.92%.

relationships between the intention to purchase and the value of mixed or conflicting emotions involved were evaluated. The three branded products considered were positioned in quadrants of the consumers' perceptual map having different connotations of emotive constructs. Purchase experiences of the brands were expressed on the basis of Positive and Negative emotions (Laros & Steenkamp, 2005). Furthermore, results showed the existence of a different and special relationship between the intention to purchase and the valences of the emotions involved.

The present research links emotions and credibility (Maathuis, *et al.*, 2004), by integrating recent findings concerning brand credibility (Erdem & Swait, 2004; Wang & Yang, 2010, 2011) with those on conflicting emotions evoked by products (Mick & Fournier, 1998; Ruth, Brunel, & Otnes, 2002; Williams & Aaker, 2002; Laros & Steenkamp, 2005; Chitturi, Raghunathan, & Mahajan, 2007; Penz & Hogg, 2011). If the goal of companies is to create

TABLE 5
INFLUENCE OF NEGATIVE AND POSITIVE EMOTION RATINGS FROM THREE TAXONOMIES
(IZARD; PLUTCHIK; O'SHAUGHNESSY AND O'SHAUGHNESSY) FOR THE EXPERTISE-PRODUCT

Variable	B	Std. Error	Beta	t
Constant	64.61	10.41		6.21†
Positive Emotions	-0.01	0.008	-0.23	-1.73†
Negative Emotions	-0.02	0.007	-0.47	-3.49†

Note.—N = 80. The dependent variable was Intention to purchase ($R = .55$; $R^2 = .31$; Adj. $R^2 = .27$; standard error = 9.91; $F_{2,79} = 8.59$; $p < .01$). † $p < .01$.

TABLE 6
FACTOR LOADINGS OF EMOTION RATINGS FROM THREE TAXONOMIES (IZARD; PLUTCHIK;
O'SHAUGHNESSY AND O'SHAUGHNESSY) CONCERNING THE TRUSTWORTHINESS-PRODUCT

Positive Emotion	Factor Loading	Communalities	Negative Emotion	Factor Loading	Communalities
Love	.90	0.80	Disgust	.81	0.66
Interest	.84	0.70	Shame	.78	0.61
Pride	.84	0.70	Embarrassment	.77	0.59
Integrity	.81	0.65	Anger	.67	0.47
Joy	.78	0.60	Sadness	.65	0.45
Hope	.77	0.59	Contempt	.65	0.42
Expectancy	.76	0.59	Fear	.61	0.47
Surprise	.75	0.69	Remorse	.60	0.37
Acceptance	.74	0.55	Guilt	.59	0.36

Note.—Eigenvalues of the factors after rotation: Positive Emotions = 32.93, Negative Emotions = 23.52; Total amount of variance explained = 56.45%.

a strong relationship between the consumer and the brand, the core component of this relationship is credibility. We focused on the emotional aspect of brand credibility, i.e., the emotional responses associated with the believability of brand signals as a result of their main component of Attractiveness, Expertise, and Trustworthiness evaluated in terms of evoked emotional responses. The consideration of brand emotional credibility in the positioning strategies of companies may allow them to discover how people perceive their products emotionally and, simultaneously, identify unconscious elements that influence consumption choices. The relevance of consumers' emotional engagement indicates the transition from the unique selling proposition, which underlines the functional benefits of a product, to an emotional selling proposition, which emphasizes the relevance of the unique attributes of a brand (cf., Aitchison, 1999).

This research has also shown that brand emotional credibility may be important to the intention to purchase branded products with vary-

TABLE 7
INFLUENCE OF NEGATIVE AND POSITIVE EMOTION RATINGS FROM THREE TAXONOMIES (IZARD;
PLUTCHIK; O'SHAUGHNESSY AND O'SHAUGHNESSY) FOR THE TRUSTWORTHINESS-PRODUCT

Variable	B	Std. Error	Beta	t
Constant	28.76	0.84		34.20†
Positive Emotions	7.98	0.87	0.61	9.14†
Negative Emotions	-7.90	0.90	-0.59	-8.83†

Note.— $N = 80$. The dependent variable was Intention to purchase ($R = .81$, $R^2 = .66$, $Adj. R^2 = .65$, standard error = 7.52, $F_{2,79} = 73.30$, $p < .01$). † $p < .01$.

ing credibility, confirming that the simultaneous importance of positive and negative emotions is intense and conflicting, as well as natural and commonplace in the lives of individuals (Williams & Aaker, 2002; Aviezer, *et al.*, 2012). Consumption opportunities can be marked by antithetical and opposing emotive experiences and engender the so-called consumer ambivalence, defined as the occurrence of multiple positive and/or negative emotions in a single purchase occasion (Otnes, Lowrey, & Shrum, 1997). An element of contradiction is already implicit in the act of consumption, which, in itself, involves the conflict between the pleasure of owning an object and a decrease in the amount of money available to a consumer (Rook, 1985). The intention to purchase or not to purchase branded products may result from different and opposing emotive states; this situation is evident not only in the behaviors of impulsive or compulsive buying (Ramanathan & Williams, 2007; Saraneva & Sääksjärvi, 2008), but also throughout all the stages of the buying decisions (Ruth, Brunel, & Otnes, 2002; Williams & Aaker, 2002; Penz & Hogg, 2011).

Elements that contribute to the formation of preferences may not be only positive emotions but also negative ones since consumers sometimes do not select products that they mostly prefer but rather reject those that they most dislike, which is consistent with the refusal of other individuals' tastes or consumption lifestyles (Bourdieu, 1984). For example, Attractiveness-products may be preferred as a consequence of feelings of Love or Interest toward a fashion brand or of a feeling of Dislike for other ones. A consumer may feel Joy and Surprise to find that a fashionable two-piece set is on sale, but also Sadness to realize that her size is unavailable, or Anger if salespersons are not polite or careful in selling it. In some other consumption circumstances, consumers tend to decide on rational aspects and are predominantly directed to deal with potentially negative outcomes (Penz & Hogg, 2011). Assessing Expertise-products, the purchase of a medicated toothpaste, for example, may be not encouraged by the Surprise resulting from any promotional offers or the cheapness of the product, but hampered by the Fear related to possible ineffectiveness. Furthermore, consumers can experience a broad variety of emotions in response to occurrences that are coherent with one goal and impede another (Weigert, 1991). For Trustworthiness-products, the purchase of a mobile phone can be encouraged by the Interest in having an innovative product, but discouraged in times of widespread economic crisis by the Embarrassment towards those with less money, even though a potential buyer has the funds to purchase luxury or expensive goods.

This finding has implications for both marketing managers and retailers. From the manager's standpoint, the findings emphasize the importance of creating and improving communication strategies, balancing

emotional messages in relation to the brand credibility profile. Retailers may improve store atmospherics—layout and design variables, point-of-purchase and decoration variables, human variables, and so on—with the aim of managing consumers' positive and negative emotions in relation to the main component of credibility.

To overcome limitations related to the consideration of self-reported questionnaires and convenience sampling, this research could be replicated in everyday consumer situations. Instead of a traditional measure of brand credibility (Guido, Prete, & Sammarco, 2010), the consumer's cognitive and motivational processes influencing credibility could be evaluated. Results suggest that the knowledge of mixed emotions evoked by different categories of branded products with varying credibility (brand emotional credibility) may allow identification of emotion benefits associated with different products to match consumers' desires and expectations.

REFERENCES

- AITCHISON, J. (1999) *Cutting edge advertising*. Singapore: Prentice Hall.
- AVIEZER, H., TROPE, J., & TODOROV, A. (2012) Body cues, not facial expressions, discriminate between positive and negative emotions. *Science*, 338, 1225-1229.
- BAGOZZI, R. P., GOPINATH, M., & NYER, P. U. (1999) The role of emotions in marketing. *Academy of Marketing Science Journal*, 27, 184-202.
- BOURDIEU, P. (1984) *Distinction: a social critique of the judgment of taste*. London, UK: Routledge.
- CACIOPPO, J., GARDNER, L., & BERNTSON, G. (1997) Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1), 3-25.
- CHITTURI, R., RAGHUNATHAN, R., & MAHAJAN, V. (2007) Form versus function: how the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences. *Journal of Consumer Research*, 44, 702-714.
- COHEN, J. B., & ARENI, C. S. (1991) Affect and consumer behavior. In T. S. Robertson & H. H. Kassarian (Eds.), *Handbook of consumer behavior*. Englewood Cliffs, NJ: Prentice Hall. Pp. 188-240.
- DAMASIO, A. (1999) *The feeling of what happens: body, emotion in the making of consciousness*. New York: Harcourt Brace.
- ERDEM, T., & SWAIT, J. (2004) Brand credibility, brand consideration, and choice. *Journal of Consumer Research*, 31(1), 191-198.
- FISHBEIN, M., & AJZEN, I. (1975) *Belief, attitude, intention, and behavior: an introduction to theory and research*. Reading, MA: Addison-Wesley.
- GORDON, R. M. (1987) *The structure of emotions*. Cambridge, UK: Cambridge University Press.
- GUIDO, G., & PELUSO, A. M. (2006) Effects of baby-faced endorsers on credibility, affect, and purchase intention. *European Advances in Consumer Research*, 7, 442-443.
- GUIDO, G., & PELUSO, A. M. (2009) When are baby-faced endorsers appropriate? Testing effects on credibility and purchase intention. *Journal of Current Issues and Research in Advertising*, 31(2), 67-74.

- GUIDO, G., & PELUSO, A. M. (2011) Beardedness in advertising: effects on endorsers' credibility and purchase intention. *Journal of Marketing Communications*, 17(1), 37-49.
- GUIDO, G., PRETE, M. I., & SAMMARCO, L. (2010) Prototypes versus examples: a new model of online credibility for commercial websites. *Journal of Targeting, Measurement and Analysis for Marketing*, 18(1), 3-16.
- HALEY, R. I. (1985) *Developing effective communication strategy: a benefit segmentation approach*. New York: Wiley.
- HIGGINS, E. T. (1997) Beyond pleasure and pain. *American Psychologist*, 52(12), 1280-1300.
- HIRSCHMAN, E., & STERN, B. (1999) The role of emotion in consumer research. *Advances in Consumer Research*, 26, 4-11.
- HOVLAND, C. I., IRVING, K. J., & KELLEY, H. H. (1953) *Communication and persuasion*. New Haven, CT: Yale Univer. Press.
- IZARD, C. E. (1977) *Human emotions*. New York: Plenum.
- LAROS, F., & STEENKAMP, J. B. (2005) Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research*, 58, 1437-1445.
- LARSEN, J., MCGRAW, A., & CACIOPPO, J. (2001) Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, 81, 684-696.
- LEKNES, S., & TRACEY, I. (2008) Pleasure and pain: masters of mankind. In M. Kringelbach & K. Berridge (Eds.), *Pleasures of the brain*. Oxford, UK: Oxford Univers. Press.
- LI, Y., WANG, X., & YANG, Z. (2011) The effect of corporate-brand credibility, perceived corporate-brand origin, and self-image congruence on purchase intention: evidence from China's auto industry. *Journal of Global Marketing*, 24, 58-68.
- MAATHUIS, O., RODENBURG, J., & SIKKEL, D. (2004) Credibility, emotion and reason? *Corporate Reputation Review*, 4, 333-345.
- MCGUIRE, W. J. (1985) Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. Vol. 2. New York: Random House. Pp. 233-234.
- MICK, D., & FOURNIER, S. (1998) Paradoxes of technology: consumer cognizance, emotions, and coping strategies. *Journal of Consumer Research*, 24, 123-143.
- OHANIAN, R. (1990) Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19(3), 39-52.
- O'SHAUGHNESSY, J., & O'SHAUGHNESSY, N. J. (1999) *The marketing power of emotion*. The Judge Institute of Management Studies, Univer. of Cambridge, Cambridge, UK.
- OTNES, C., LOWREY, T. M., & SHRUM, L. J. (1997) Toward an understanding of consumer ambivalence. *Journal of Consumer Research*, 24(1), 80-93.
- PENZ, E., & HOGG, M. K. (2011) The role of mixed emotions in consumer behavior. *European Journal of Marketing*, 45(1-2), 104-132.
- PLUTCHIK, R. (1980) *Emotion: a psycho-evolutionary synthesis*. New York: Harper & Row.
- PRETE, M. I. (2007) M-politics: credibility and effectiveness of mobile political communications. *Journal of Targeting, Measurement and Analysis for Marketing*, 16(1), 48-56.
- RAMANATHAN, S., & WILLIAMS, P. (2007) Influence and delayed emotional consequences of indulgence: the moderating influence of personality type on mixed emotions. *Journal of Consumer Research*, 34(2), 212-223.
- ROOK, D. W. (1985) The ritual dimension of consumer behavior. *Journal of Consumer Research*, 12 (December), 251-264.

- RUTH, J. A. (2001) Promoting a brand's emotion benefits: the influence of emotion categorization processes on consumer evaluations. *Journal of Consumer Psychology*, 11(2), 99-113.
- RUTH, J. A., BRUNEL, F. F., & OTNES, C. C. (2002) Linking thoughts to feelings: investigating cognitive appraisals and consumption emotions in a mixed emotions context. *Journal of the Academy of Marketing Science*, 30(1), 44-58.
- SARANEVA, A., & SÄÄKSJÄRVI, M. (2008) Young compulsive buyers and the emotional roller-coaster in shopping. *Young Consumers: Insight and Ideas for Responsible Marketers*, 9(2), 75-79.
- WANG, X., & YANG, Z. (2010) The effect of brand credibility on consumers' brand purchase intention in emerging economies: the moderating role of brand awareness and brand image. *Journal of Global Marketing*, 23, 177-188.
- WANG, X., & YANG, Z. (2011) The impact of brand credibility and brand personality on purchase intention: an empirical study in China. *Advances in International Marketing*, 21, 137-153.
- WEIGERT, A. J. (1991) *Mixed emotions: certain steps toward understanding ambivalence*. Albany, NY: State Univer. of New York Press.
- WILLIAMS, P., & AAKER, J. (2002) Can mixed emotions peacefully co-exist? *Journal of Consumer Research*, 28(4), 636-649.
- WIND, Y. J. (1990) Positioning analysis and strategy. In G. Day, B. Weitz, & R. Wensley (Eds.), *The interface of marketing strategy*. Greenwich, CT: JAI Press. Pp. 387-412.

Accepted September 17, 2013.

PREDICTING GAME-ATTENDING BEHAVIOR IN AMATEUR ATHLETES: THE MODERATING ROLE OF INTENTION STABILITY¹

WAN CHEN LU

*Department of Tourism and Leisure Management
Taoyuan Innovation Institute of Technology, Taiwan*

CHIH-FU CHENG

*Department of Physical Education
National Taiwan Normal University, Taiwan*

LUNG HUNG CHEN

*Department of Recreation and Leisure Industry Management
National Taiwan Sport University, Taiwan*

Summary.—The theory of planned behavior is a well-established theory in predicting human behavior. However, there is evidence of an inconsistent relationship between intention and behavior. Therefore, the purpose of the current study is to further investigate the gap between intention and behavior. The study proposes intention stability as the moderator. Participants ($N = 154$, M age = 23 yr., $SD = 6.7$) were recruited from Internet volleyball forums and local volleyball courts in Taiwan. Multiple hierarchical regression was used to analyze the data. The results indicated that perceived behavioral control significantly predicted game-attending behavior through intention. However, attitude and subjective norms did not significantly predict behavioral intention. In addition, intention stability moderated the relationship between intention and behavior and indicated the relationship between intention and behavior was strong when intention stability was high. On the contrary, when intention stability was low, the relationship between intention and behavior was weak. Implications and applications are discussed.

Non-elite mass participation in sporting events has increased worldwide. For example, in the 2010 London marathon, less than 0.5% of the 37,000 participants were elite athletes, while 20% were club runners (Coleman & Ramchandani, 2010). Sporting events for mass participation are widespread and growing rapidly, bringing great economic value to the sporting industry and host cities. To date, more and more studies have focused on the benefits of mass participation in sports events, respective of the great marketing opportunity (Agrusa, Lema, Kim, & Botto, 2009; Coleman & Ramchandani, 2010). However, academic literature relating to sport attendance has primarily focused on spectators and fans, such as

¹Address correspondence to Lung Hung Chen, Department of Recreation and Leisure Industry Management, National Taiwan Sport University, Taiwan, No. 250, Wen Hua 1st Road Kueishan, Taoyuan County, Taiwan or e-mail (tjudragon@gmail.com).

spectators' motivation or satisfaction (Trail & James, 2001). In comparison to spectators, research discussing the participation of amateur athletes is less developed, which needs to be addressed and further understood.

This paper focuses on the game attending behavior in amateur volleyball athletes. Volleyball is one of the most popular sports in Asia, and Taiwan is no exception. More than 50 volleyball competitions are held every year in Taiwan. Most of the competitions are at the local or grassroots level and appeal to participants who are, for the most part, non-elite players. For example, one of the biggest volleyball events held in Taiwan is the Yungshin Cup Volleyball Championships. Every year it attracts more than 300 teams and 3,000 players (YungShin Group, 2012), and brings additional economic benefits for the host city.

In this regard, the researcher adapted the theory of planned behavior as a framework for the current study. For the past decades, the theory of planned behavior has been widely applied in sports. A number of studies and meta-analytic reviews support the utility of the theory in predicting exercise behavior (Armitage & Conner, 2001; Hagger, Chatzisarantis, & Biddle, 2002; Kerner & Kurrant, 2003; Downs & Hausenblas, 2005; McEachan, Conner, Taylor, & Lawton, 2011). Furthermore, some studies have applied this theory to sports competition, in areas such as training adherence (Palmer, Burwitz, Dyer, & Spray, 2005; Anderson & Lavallee, 2008) and training participation (Theodorakis, Goudas, Bagiatis, & Doganis, 1991). However, little is known about the amateur athletes' decision processes involved in game participation (Kaplanidou & Gibson, 2010). Thus, the current study tried to investigate the amateur athletes' game-attending behavior as players by applying the theory of planned behavior in an attempt to improve understanding of sports participation behavior.

The theory of planned behavior assumes that intention is the proximal determinant of that behavior. Intention outlines an individual's motivation to execute a behavior, and how hard people are willing to work (Ajzen, 1991). It also served as the mediator between the three constructs (attitude, subjective norms, and perceived behavioral control) and behavior. Attitude is one's positive or negative evaluation of performing a certain behavior. Subjective norms are one's perceptions of whether others approve or disapprove of the behavior. Perceived behavioral control reflects one's perceived capability to execute the behavior. When behavior is not under complete volitional control, perceived behavioral control can improve a prediction toward behavior (Madden, Ellen, & Ajzen, 1992). As the three factors increase, the intention toward certain behavior may ultimately rise.

Since playing in a volleyball game is complicated and has multiple determinants, such as cognitive, effective, and environmental factors, the

theory of planned behavior may provide a sound theoretical framework for exploring such behaviors. For instance, an athlete may intend to participate in a volleyball game because it is his or her hobby (attitude), or as a result of encouragement of coaches and teammates (subjective norms). In addition, players may be willing to join the game because of having excellent skills (perceived behavioral control). Meta-analytic reviews provide strong evidence for the use of the theory of planned behavior in predicting behavior. For example, Armitage and Conner (2001) reported that the theory of planned behavior accounted for 27% and 39% of variance in behavior and intention, respectively. Godin and Kok (1996) reported that the theory of planned behavior accounted for 41% of the variance in intentions and 34% of the variance in behaviors for a domain of health behaviors.

Despite the predictive power of the theory of planned behavior, a huge gap still exists between an individual's intentions and behavior (Sheeran & Orbell, 1999; Sheeran & Abraham, 2003). For instance, a review of health behaviors reported that 47% of participants with high intentions ultimately failed to perform their intended activities (Sheeran, 2002). A meta-analytic study of the intention-behavior relation proposed that intention explains 28% of the variance in actual behavior (Sheeran, 2002). In addition, Hagger, *et al.* (2002) conducted a meta-analysis of 79 studies and found that intentions only account for 27.4% of the variance in physical activity behavior.

Intentions may change due to new information or unforeseen obstacles and result in reduced predictive power toward behavior (Conner & Godin, 2007). Ajzen (1991) suggested that extra predictors could be recognized in the theory of planned behavior if they accounted for appreciable variance. One approach to improve the predictive power of the theory is to examine the moderating variables. Past studies had identified several moderators between intention and behavior, including intention certainty, past behavior, self-schema, anticipated regret, attitudinal control, and temporal stability (Sheeran & Abraham, 2003). Among above variables, temporal stability may serve as one of the strongest moderators to improve the consistency between intentions and behavior (Cooke & Sheeran, 2004).

Sheeran, Orbell, and Trafimow (1999) defined temporal stability as "the extent to which an attitude remains unchanged over time regardless of whether or not it is challenged" (p. 725). At the theoretical level, Ajzen (1991) stated that "the second condition for accurate behavioral prediction is that intentions and perceived behavioral control must remain stable in the interval between their assessment and observation of the behavior" (p. 185). As support for Ajzen's hypothesis, intention stability could be served as a boundary of a relation between intention and behavior. People

with stable intentions are less likely to change intention prior to behavior compared to those with unstable intentions (Sheeran & Abraham, 2003). Meta-analytic research has shown that temporal stability is a statistically significant moderator between intention and behavior (Cooke & Sheeran, 2004; Conner & Godin, 2007). Based on the above discussion, it seems that intention stability could be adopted as a moderating role in the relationship between intention and behavior.

The goal of this study was to further understand determinants of game-attending behavior and improve the predictive power of the theory of planned behavior by examining the moderating role of temporal stability. Although this moderating relationship has been extensively examined in the health and exercise domains, there have been few studies (e.g., Kaplanidou & Gibson, 2010) to investigate predictable variables on game-attending behavior in amateur athletes. Since amateur sport is the foundation of elite and professional sport, and provides a great marketing opportunity for the sports industry (e.g., sportswear and merchandise) to target its consumers, it is important to understand the critical factors of amateur athletes to maintain their game-attending behavior.

METHOD

Participants

A sample of 464 volleyball participants was recruited from Internet volleyball forum members ($n=247$, 53.2%) and local volleyball courts at two public sport parks and two university campuses ($n=217$, 46.8%). Participants joined this study voluntarily and their confidentiality was guaranteed. All 464 participants completed the initial questionnaire and 271 participants (58.4%) completed both the initial and the follow-up questionnaire. Participants who completed both the initial and the follow-up questionnaire ($n=154$; 33.2%) were successfully tracked to complete the final questionnaire. Of the participants, 45.5% were male, 37.1% were under 21 years old, and 52.5% were ages 21 to 30 years old.

Of all three longitudinal waves, 41.6% of participants were lost at Wave 2, and 25.2% were lost at Wave 3. Participants lost at Wave 2 did not return for Wave 3. This resulted in a loss of 193 participants in Wave 2 and 117 participants in Wave 3. Each subsample exhibited a random missing-data pattern. Consequently, listwise deletion was adopted as an unbiased technique under the MCAR condition. In addition, due to the high attrition rate within the data collection with three waves, a Student *t* test was analyzed to examine whether there are differences between the study completers and drop-outs. The results indicated that there are no significant differences on age, gender, site (Internet and on-the-spot delivery), attitude, and subjective norms. However, there are differences in intention

(for completers, $M=5.54$, and drop-outs $M=5.31$, $p<.05$) and perceived behavioral control (for completers, $M=6.11$, and drop-outs $M=5.67$, $p<.05$). The results indicated that individuals who completed the study are higher in intention and perceived behavioral control than drop-outs. This is probably the reason why these people accomplished the investigation at all three time points.

Measures

The model constructs and measures were adopted from Ajzen's (2002) standard direct measure. The instrument was translated from English to Chinese, and back-translation was undertaken to guarantee that the translation did not alter the intent of the questionnaire. A bilingual assistant professor with expertise in sport psychology translated the scales from English to Chinese. The translated measures were then translated back to English by another bilingual expert who was blind to the original questionnaires. Finally, both the original and back-translated measures were also compared by the authors until the translated questionnaire was satisfactory.

Attitude toward the behavior was defined by Ajzen's (2002) standard direct measure. Attitude was defined as a person's overall evaluation of attending volleyball games as a player in half a year. Five items with semantic differential pairs (e.g., harmful–beneficial, pleasant–unpleasant, good–bad) were used to measure attitude. Participants rated these items with 7-point bipolar adjective scales ranging from 1 (e.g., unpleasant) to 7 (e.g., pleasant). A sample item is, "For me, attending a volleyball game as a player in half a year is ____ ('unpleasant' to 'pleasant')." The internal consistency score of these five items was excellent ($\alpha = .94$).

Subjective norms were measured by 7-point adjective bipolar scales from Ajzen's (2002) standard direct measure. Subjective norms were defined as one's perceptions of pressure coming from significant others who think the player ought or ought not to participate in volleyball games in half a year. Four items with semantic differential pairs (e.g., should–should not, approve–disapprove) were used to measure subjective norms and ranged from 1: Disapprove to 7: Approve. This is a sample item: "Most people who are important to me think that I ____ ('should' to 'should not') attend a volleyball game as a player in half a year." The internal consistency score of these four items was good ($\alpha = .82$).

Perceived behavioral control was measured using 7-point bipolar adjective scales suggested by Ajzen's (2002) standard direct measure. It was defined as a person's perception of ease or difficulty in participating in volleyball games in half a year. Four items with the semantic differential pairs (e.g., impossible–possible, no control–complete control) were used to measure perceived behavioral control with anchors 1: Impossible and 7:

Possible. A sample item is, "For me to attend volleyball games as a player in half a year would be ____ ('impossible' to 'possible')." The internal consistency score of the four items was good ($\alpha = .80$).

Intention was measured using 7-point bipolar adjective scales suggested by Ajzen's (2002) standard direct measure. Intention was assessed in terms of a participant's subjective probability judgments that he or she would attend volleyball games in half a year. Three items with the semantic differential pairs (e.g., definitely false–definitely true) were used to measure intention with anchors 1: Definitely false and 7: Definitely true. A sample item is, "I will try to attend volleyball games as a player in half a year, and it would be ____ ('definitely false' to 'definitely true')." The internal consistency score of the Time 1 intention of these three items was excellent ($\alpha = .92$), and so was Time 2 intention ($\alpha = .93$).

Intention stability has been measured with a number of indices reported in the literature. Based on previous research (Conner, Sheeran, Norman, & Armitage, 2000), the researcher used the sum of absolute differences between intention items at two time points to serve as the indices of intention stability. To ensure that high scores on the stability measure represented higher stability, the scores were reversed (new value = 7 – original value).

Behavior was measured six months after the initial questionnaire with a self-reported behavioral measure to assess the game-attending behavior of participants as players. This variable was measured in response to the question: "How many volleyball games have you attended as a player during half a year?" Researchers made a list of domestic volleyball games held during the last six months, and asked participants to report those they had attended (played in) and counted this number as their behavioral index. The researcher adopted this item as the index of self-report behavior and regarded the data as ratio values in the current study.

Procedure

The study design included data collection at three time-points. The data of attitude, subjective norms, perceived behavioral control, and intention were collected by an initial questionnaire through the Internet and on-the-spot delivery. The researcher chose three local volleyball courts (one in a sports center and the others on school campuses) to deliver on-the-spot questionnaires, and contacted two on-line volleyball groups' members as the e-survey participants. Participants were assured of anonymity and informed that their participation was entirely voluntary. This study also informed participants that they could drop out of this project without excuse. After two months of accomplishing the initial questionnaire, the intention stability data was collected by a follow-up

questionnaire. The final questionnaire was delivered to collect the behavioral data after six months from the initial measure. E-mail and telephone interviews were used to gather the data of both the follow-up and final questionnaires.

Analysis

To determine whether the expected factor structure could be supported, four constructs including attitude, subjective norms, perceived behavioral control and intention were conducted for factor analysis. Based on Tabachnick and Fidell's (2007) recommendation, this research adopted principal axis factoring followed by Promax oblique rotations to clearly separate factors of TPB direct measures (attitude, subjective norms, perceived behavioral control, and intention). The results showed that the Kaiser-Meyer-Olkin (KMO) value of TPB direct measures was 0.85 and Bartlett's sphericity was significant ($p < .001$). The results showed that no items were removed from original concepts after rotation. Four factors were confirmed, with proportions of variance explained, 41.34%, 15.32%, 12.45%, and 7.15%, respectively, accounting for 76.26% of the total variance.

TABLE 1
FACTOR ANALYSIS FOR THEORY OF PLANNED BEHAVIOR ITEMS

Scale	Item	Factor			
		1	2	3	4
Attitude	a3	.96	-.09	-.003	.07
	a4	.93	-.001	-.05	.02
	a2	.91	-.02	-.09	.12
	a5	.87	.01	.06	-.09
	a1	.58	.18	.22	.04
Subjective Norms	nm3	-.06	.92	-.19	.19
	nm4	-.10	.85	-.12	.20
	nm2	.16	.80	.11	-.23
	nm1	.01	.79	.24	-.18
Intention	bi1	.00	.04	.92	.02
	bi2	.08	-.07	.90	.02
	bi3	.01	-.02	.89	.07
Perceived Behavioral Control	pc2	-.25	.007	.50	.49
	pc3	.04	.08	-.06	.85
	pc4	.07	-.09	.13	.72
	pc1	.16	.02	.03	.71

RESULTS

Descriptive Analyses

Table 2 presents the descriptive statistics (means, standard deviations, and correlations) for the study variables. The results indicated that attitude, perceived behavioral control, and behavioral intention were positively related with each other. Subjective norms were positively associated with perceived behavioral control and intention. Intention was positively related to intention stability and behavior. In addition, intention stability and behavior were positively related to each other.

TABLE 2
DESCRIPTIVE STATISTICS AND CORRELATIONS AMONG VARIABLES

Variable	M	SD	1	2	3	4	5
Attitude	6.25	0.93					
Subjective Norms	4.59	1.15	.10				
Perceived Behavioral Control	5.54	0.97	.29‡	.17*			
Intention	6.11	1.15	.26†	.17*	.53		
Intention Stability	5.95	1.04	.13	-.10	.34†	.30‡	
Behavior	3.46	3.05	-.01	.06	.27†	.32‡	.25†

* $p < .05$, † $p < .01$, ‡ $p < .001$.

Mediation Analyses

The researchers adopted multiple hierarchical regression analysis to examine the mediating role of behavioral intention on the relationship between the independent variables (attitude, subjective norms, and perceived behavioral control) and behavior. Following Baron and Kenny's (1986) procedure, the first step was to investigate the relationship between the three independent variables (attitude, subjective norms, and perceived behavioral control) and behavior (see Table 3; Model 2; behavior $F_{3,150} = 4.42$, $p < .01$). The results indicated that only perceived behavioral control significantly affected behavior. Thus, this study partly supported Baron and Kenny's first assumption. In the second step, the researcher compared the relationship between the three independent variables and the mediating variable intention. As shown in Table 3, Model 1, the overall model was significant ($F_{3,150} = 20.93$, $p < .001$). The results revealed that only perceived behavioral control significantly predicted intention.

In the third step, the relationships among the three independent variables, intention, and behavior were further tested. As shown in Table 3, Model 3, the overall model was significant (behavior $F_{4,149} = 5.70$, $p < .001$).

However, only intention significantly predicted behavior. Although the coefficients of the three independent variables decreased from Model 2 to Model 3, the results showed that intention could only fully mediate the relationship between perceived behavioral control and behavior, because attitude and subjective norms were not statistically significant for behavior in Step 3. The results are shown in Table 3.

TABLE 3
MULTIPLE REGRESSIONS OF ATTITUDE, SUBJECTIVE NORMS, AND PERCEIVED
BEHAVIORAL CONTROL ON INTENTION AND BEHAVIOR

Variable	Intention	Behavior	
	Model 1	Model 2	Model 3
Attitude	.12	-.09	-.12
Subjective Norms	.08	.02	.00
Perceived Behavioral Control	.48*	.29*	.17
Intention			.27*
R^2	.28	.06	.11
F	20.93‡	4.42†	5.70‡
df	3,150	3,150	4,149

* $p < .05$, † $p < .01$, ‡ $p < .001$.

Moderation Analyses

The researchers used moderated hierarchical regression analysis to test the moderating effect of intention stability between intention and behavior. As recommended by Aiken and West (1991), there are three steps in hierarchical regression analysis. In the first step intention was entered, and the intention stability as the moderator was entered in the second step. In the third step, the interaction term between intention and intention stability was entered. Prior to all analyses, variables were standardized to reduce potential multicollinearity (Aiken & West, 1991).

The results of the moderated hierarchical regression analysis are shown in Table 4. Significant interaction effects were found. A moderating effect of intention stability was indicated by a significant ($p < .05$) increase in R^2 when the interaction term was entered. In the analyses, the product of Intention \times Intention stability accounted for significant variance in the model predicting behavior. However, the interaction effects as the amount of variance reported above were small (R^2 change increment $\leq .04$). For illustrative purposes, a slope was computed using a procedure suggested

TABLE 4

MODERATED HIERARCHICAL REGRESSION ANALYSIS RESULTS FOR INTENTION-BEHAVIOR RELATIONSHIP

		β^b	Adj R^2	ΔR^2
Step 1	Intention	.29‡	.10	.11‡
	$F_{1,152} = 17.81‡$			
Step 2	Intention stability	.14	.12	.02*
	$F_{2,151} = 11.28‡$			
Step 3	Intention \times Intention stability	.18*	.15	.03*
	$F_{3,150} = 9.76‡$			

Note.—^bStandardized β s from the final step. * $p < .05$. ‡ $p < .001$.

by Stone and Hollenbeck (1989), one at one standard deviation below the mean (low) and one at one standard deviation above the mean (high). Figure 1 demonstrates that intention stability will moderate the relationship between intention and behavior, and indicates that the relation is stronger when intention stability is high.

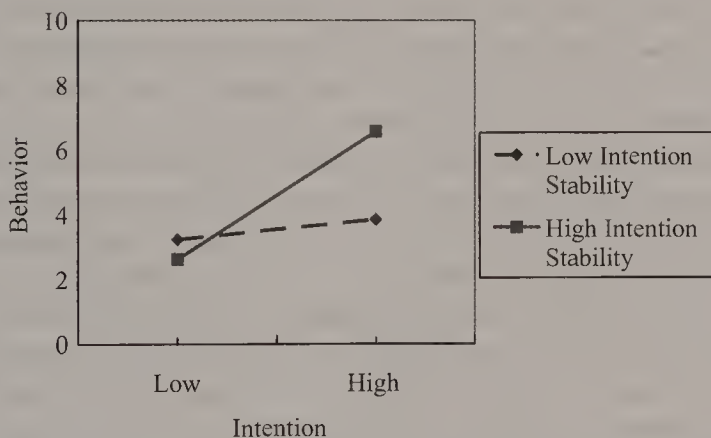


FIG. 1. Plot of the interaction of intention and intention stability on behavior.

DISCUSSION

The purpose of this study was to investigate game-attending behavior adopting the theory of planned behavior. Moreover, the researchers aimed to determine the moderating role of temporal stability between intention and behavior. As predicted by the theory, perceived behavioral control statistically predicted game-attending behavior through intention. Attitude and subjective norms, however, did not significantly predict intention and behavior. The results partially support the theory's predictions.

The present study revealed that perceived behavioral control related to behavior through intention. In the theory of planned behavior, per-

ceived behavioral control presents one's confidence to perform a certain behavior (Ajzen, 2002). As mentioned, joining a volleyball game is not an easy task to achieve. Participants may overcome those barriers (e.g., time management, teammate recruitment, tactics arrangement) by thinking deeply and planning carefully. Such activity is not only largely under volitional choice (Hrubes, Ajzen, & Daigle, 2001), but also executed with firm intention.

In the current study, attitude and subjective norms failed to statistically predict intention. The results did not correspond with past studies. Reviews of the theory of planned behavior have concluded that the attitude is the dominant factor in predicting intention in the domain of physical activity (Hagger, Chatzisarantis, & Biddle, 2001). In the current study, people with a positive attitude may not have intended to join a volleyball game. Since attending volleyball competitions is a task-oriented behavior, people with a positive attitude may not have enough to trigger their intention to perform.

Furthermore, meta-analytic studies have shown that the subjective norms-intention relation is significantly weaker than the attitude-intention or perceived behavioral control-intention relations (Armitage & Conner, 2001). Current findings agreed with previous studies and showed that subjective norms failed to relate to intention. According to Ajzen's (1991) definition, subjective norms function as a global perception of social pressure to comply with others' wishes. However, social pressure is rarely so direct or explicit (Armitage & Conner, 2001), so people may not perceive the existence of social pressure. Besides, the results of the hierarchical regression analysis revealed that intention stability moderated the effect of intention on behavior. The present study supported past research (Sheeran, *et al.*, 1999; Conner, *et al.*, 2000; Conner & Godin, 2007), and showed that when intention is stable, it becomes a strong predictor of behavior, whereas unstable intention was weakly related to behavior.

The main contribution of the current study is to validate the steadfast effect of the intention stability on the moderating role between intention and behavior. Even across different task contexts (from exercise activity to game-attending behavior), the relationship still exists. As Conner, *et al.* (2000) mentioned, stable intention is more likely to translate into subsequent action. The present study also revealed that people who have stable intentions to participate in a volleyball contest are more likely to perform it. Moreover, the results provide a different pattern for the interaction between intention and intention stability. For an individual with low intention of playing a volleyball game, intention stability does not have a great effect on behavior. On the other hand, with people showing high intention to play volleyball games, intention stability becomes a critical

factor in the frequency of performance. Thus, intention stability ensures that high intenders are more likely to perform the subsequent behavior than low intenders.

The finding is important because past research in examining game-attending behavior often gathered data merely on intention, and rarely discussed the determining factors between intention and behavior (Kaplanidou & Gibson, 2010). These results provide a clear description of how intention stability influences the relationship between intention and behavior.

Theoretically, the significance of this study is its contribution toward knowledge development, especially since past research had seldom examined the intention-behavior gap in game-attending studies (Cunningham & Kwon, 2003; Norman, Clark, & Walker, 2005; Lu, Lin, & Cheng, 2011). Individuals with equivalent intention might differ in their subsequent behaviors. The present study showed that intention stability could serve as a crucial moderator of people's intention toward action (Sheeran & Abraham, 2003).

Secondly, past researchers have focused on sport events, and often gathered data at the same time and via the same method, which may inflate the relationships between variables and threaten internal validity (Lindell & Whitney, 2001). To overcome this shortcoming, the present researchers adopted a three-wave design and mixed methods (e-mail, phone, on-the-spot survey, and e-survey) to collect data. Additionally, compared to past studies focusing merely on college students and school athletes, this study broadens the varieties of game participants by gathering data at three local volleyball courts and two on-line volleyball groups. The results may present a better picture of volleyball participants.

The results summarized above might provide some suggestion to investigate sports consumer behavior, especially in the context of sporting events. In the present study, the results revealed that perceived behavioral control could affect behavior through intention, which suggests that self-confidence to attend a volleyball game may influence actual behavior effectively. Thus, to promote participation confidence, the organizer could classify competition intensity by the players' performance levels. Past research also stated that people (especially women and older adults) drop out of sports because participation is too competitive to see themselves as competent (Biddle, Coalter, O'Donovan, MacBeth, Nevill, & Whitehead, 2005). By classifying competition levels, people who are looking for relaxation could participate in fun games, while others more interested in the thrill of victory could also enjoy the sense of achievement. Such a promotional strategy may increase the competitive balance between matches and further expand the population of game participants. Furthermore, the present study suggested that stable intentions were a strong predictor of behavior, while unstable inten-

tions had weak effect on behavior. Interventions such as e-mail notices of events may increase the intention stability of participants, and thus could be used to promote game-attending behavior of participants.

Several limitations should be addressed when interpreting the present results. Firstly, the main limitation of this study was the low return rate (154/464, 33.2%), and it may lower the representativeness of the population. Additionally, there were differences between the study completers and dropouts on intention and perceived behavioral control. Thus, the likelihood of being completers or dropouts may be associated with these two variables. It may imply that there was selection bias within the group, affecting results and conclusions. Future researchers should overcome this shortcoming by enhancing the technique of data collection and maintaining the sample size to avoid selection bias. Secondly, in this study, game-attending behavior was measured by self-report measures, and may result in inherent biases. Thus, future researchers are encouraged to assess behavioral data with a combination of self-reporting and objective measures (e.g., players list). Thirdly, the current study did not investigate the prior game-attending behavior of individuals, so the researcher could not control for past behavior in prospective design. As Weinstein (2007) argued, with ongoing or repeated behaviors, effects of behaviors on perceptions are likely to substantially inflate the correlation between perception and behavior (p. 3). Therefore, future investigations need to examine the influence of past actions when predicting target behaviors.

REFERENCES

- AGRUSA, J., LEMA, J. D., KIM, S. S., & BOTTO, T. (2009) The impact of consumer behavior and service perceptions of a major sport tourism event. *Asia Pacific Journal of Tourism Research*, 14(3), 267-277. DOI: 10.1080/10941660903023960
- AIKEN, L. S., & WEST, S. G. (1991) *Multiple regression: testing and interpreting interactions*. Newbury Park, CA: Sage.
- AJZEN, I. (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- AJZEN, I. (2002) Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32, 665-683.
- ANDERSON, A. G., & LAVALLEE, D. (2008) Applying the theories of reasoned action and planned behavior to athlete training adherence behavior. *Applied Psychology*, 57(2), 304-312. DOI: 10.1111/j.1464-0597.2007.00310.x
- ARMITAGE, C. J., & CONNER, M. (2001) Efficacy of the theory of planned behaviour: a meta-analytic review. *British Journal of Social Psychology*, 40, 471-499.
- BARON, R. M., & KENNY, D. A. (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- BIDDLE, S., COALTER, F., O'DONOVAN, T., MACBETH, J., NEVILL, M., & WHITEHEAD, S. (2005) *Increasing demand for sport and physical activity by girls*. Edinburgh, UK: Sportscotland.

- COLEMAN, R., & RAMCHANDANI, G. (2010) The hidden benefits of non-elite mass participation sports events: an economic perspective. *International Journal of Sports Marketing and Sponsorship*, 12(1), 24-36.
- CONNER, M., & GODIN, G. (2007) Temporal stability of behavioural intention as a moderator of intention-health behaviour relationships. *Psychology & Health*, 22(8), 875-897. DOI: 10.1080/14768320601070449
- CONNER, M., SHEERAN, P., NORMAN, P., & ARMITAGE, C. J. (2000) Temporal stability as a moderator of relationships in the theory of planned behaviour. *British Journal of Social Psychology*, 39(4), 469-493. DOI: 10.1348/014466600164598
- COOKE, R., & SHEERAN, P. (2004) Moderation of cognition-intention and cognition-behaviour relations: a meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology*, 43(2), 159-186. DOI: 10.1348/0144666041501688
- CUNNINGHAM, G. B., & KWON, H. (2003) The theory of planned behaviour and intentions to attend a sport event. *Sport Management Review*, 6(2), 127-145.
- DOWNS, D. S., & HAUSENBLAS, H. A. (2005) The theories of reasoned action and planned behavior applied to exercise: a meta-analytic update. *Journal of Physical Activity and Health*, 2(1), 76-97.
- GODIN, G., & KOK, G. (1996) The theory of planned behavior: a review of its applications to health-related behaviors. *American Journal of Health Promotion*, 11(2), 87-98. DOI: 10.4278/0890-1171-11.2.87
- HAGGER, M. S., CHATZISARANTIS, N., & BIDDLE, S. J. H. (2001) The influence of self-efficacy and past behavior on the physical activity intentions of young people. *Journal of Sports Sciences*, 19(9), 711-725.
- HAGGER, M. S., CHATZISARANTIS, N. L. D., & BIDDLE, S. J. H. (2002) A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: predictive validity and the contribution of additional variables. *Journal of Sport and Exercise Psychology*, 24, 3-32.
- HRUBES, D., AJZEN, I., & DAIGLE, J. (2001) Predicting hunting intentions and behavior: an application of the theory of planned behavior. *Leisure Sciences*, 23, 165-178.
- KAPLANIDOU, K., & GIBSON, H. J. (2010) Predicting behavioral intentions of active event sport tourists: the case of a small-scale recurring sports event. *Journal of Sport & Tourism*, 15(2), 163-179.
- KERNER, M. S., & KURRANT, A. B. (2003) Psychosocial correlates to high school girls' leisure-time physical activity: a test of the theory of planned behavior. *Perceptual & Motor Skills*, 97, 1175-1183.
- LINDELL, M. K., & WHITNEY, D. J. (2001) Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86, 114-121.
- LU, W.-C., LIN, C.-H., & CHENG, C.-F. (2011) Sports spectator behavior: a test of the theory of planned behavior. *Perceptual & Motor Skills*, 113(3), 1017-1026.
- MADDEN, T. J., ELLEN, P. S., & AJZEN, I. (1992) A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, 18, 3-9.
- MC EACHAN, R. R. C., CONNER, M., TAYLOR, N. J., & LAWTON, R. J. (2011) Prospective prediction of health-related behaviours with the theory of planned behaviour: a meta-analysis. *Health Psychology Review*, 5(2), 97-144.

- NORMAN, P., CLARK, T., & WALKER, G. (2005) The theory of planned behavior, descriptive norms, and the moderating role of group identification. *Journal of Applied Social Psychology*, 35(5), 1008-1029.
- PALMER, C. L., BURWITZ, L., DYER, A., & SPRAY, C. (2005) Endurance training in elite junior netball athletes: a test of the theory of planned behavior and a revised theory of planned behavior. *Journal of Sports Sciences*, 23, 277-288.
- SHEERAN, P. (2002) Intention-behavior relations: a conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1-36. DOI: 10.1080/14792772143000003
- SHEERAN, P., & ABRAHAM, C. (2003) Mediator of moderators: temporal stability of intention and the intention-behavior relationship. *Personality and Social Psychology Bulletin*, 29, 205-215.
- SHEERAN, P., & ORBELL, S. (1999) Augmenting the theory of planned behavior: roles for anticipated regret and descriptive norms. *Journal of Applied Social Psychology*, 29(10), 2107-2142. DOI: 10.1111/j.1559-1816.1999.tb02298.x
- SHEERAN, P., ORBELL, S., & TRAFIMOW, D. (1999) Does the temporal stability of behavioral intentions moderate intention-behavior and past behavior-future behavior relations? *Personality and Social Psychology Bulletin*, 25(6), 724-734. DOI: 10.1177/0146167299025006007
- STONE, E. F., & HOLLENBECK, J. R. (1989) Clarifying some controversial issues surrounding statistical procedures for detecting moderating variables: empirical evidence and related matters. *Journal of Applied Psychology*, 74(1), 3-10.
- TABACHNICK, B. G., & FIDELL, L. S. (2007) *Using multivariate statistics*. (5th ed.) Boston, MA: Pearson.
- THEODORAKIS, Y., GOUDAS, M., BAGLIS, K., & DOGANIS, G. (1991) Reasoned action theory and the prediction of training participation in young swimmers. *British Journal of Physical Education*, 10, 10-13.
- TRAIL, G. T., & JAMES, J. D. (2001) The motivation scale for sport consumption: assessment of the scale's psychometric properties. *Journal of Sport Behavior*, 24(1), 108-127.
- WEINSTEIN, N. D. (2007) Misleading tests of health behavior theories. *Annals of Behavioral Medicine*, 33(1), 1-10. DOI: 10.1207/s15324796abm3301_1
- YUNGSHIN GROUP. (2012) [*The history of the Yungshin Cup Volleyball Championships*]. Retrieved from <http://vbg.ysp.com.tw/vbg/Idea.aspx>. [in Chinese]

Accepted October 10, 2013.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Measures &
Statistics*



LEVELS OF RECOVERY SCALE (LORS): PSYCHOMETRIC PROPERTIES OF A NEW INSTRUMENT TO ASSESS PSYCHOTIC SYMPTOMS AND PATIENT AWARENESS¹

DONALD P. CORRIVEAU AND SHARON SOUSA

University of Massachusetts–Dartmouth

Summary.—This study developed norms and examined psychometric properties of the LORS, an instrument designed to assess symptoms of psychotic illness and patients' awareness of their symptoms. Participants were 90 patients with psychotic illnesses. The LORS was completed by trained clinicians (LORS–Clinician Scale) and patients themselves (LORS–Patient). To assess patients' self-awareness of psychotic symptoms, a derived scale (LORS–Discrepancy) was computed by subtracting the LORS–Patient Scale score from the LORS–Clinician Scale score. All scales showed good test-retest reliability. The LORS–Clinician Scale showed good concurrent validity. LORS–Patient Scale scores were significantly lower than LORS–Clinician Scale scores, suggesting that patients underestimated their psychotic symptoms. Patients with low awareness of their psychotic symptoms showed less appropriate functioning, more problems in complying with their medication regimens, and increased positive and negative symptoms of schizophrenia. Implications for developing treatments using the LORS–Discrepancy Scale were discussed.

Despite improvements in the pharmacological treatment of psychotic disorders, relapse remains a major problem for patients diagnosed with schizophrenia spectrum disorders (Masand, Roca, Turner, & Kane, 2009; Gaebel, Schreiner, Bergmans, de Arce, Rouillon, Cordes, *et al.*, 2010; Novick, Haro, Suarez, Perez, Dittmann, & Haddad, 2010). One variable thought to affect relapse is non-adherence to medication regimens (Cramer & Rosenheck, 1998; Lacro, Dunn, Dolder, Leckband, & Jeste, 2002). Another variable appears to be “lack of patient-insight” (Linden & Gode-mann, 2007; Beck, Cavelti, Kvrjic, Kleim, & Vauth, 2011). While several conceptual models have been applied to the insight construct (Chakraborty & Basu, 2010), a most common definition centers on patients' self-awareness of psychotic symptoms. While an assessment of patients' self-awareness would seem clinically relevant, Lincoln, Lüllmann, and Rief (2007) critically examined 13 published measures of insight and concluded that many of the measures did not directly assess self-awareness.

Pragmatically, a better understanding of patients' self-awareness of psychotic symptoms may aid in providing better treatment. Sousa (1998)

¹Address correspondence to Donald P. Corriveau, Ph.D., University of Massachusetts–Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747 or e-mail (DCorriveau@UMassD.edu).

developed a simplified interview and scoring system called the Levels of Recovery Scale (LORS). Notably, this scale included both positive and negative symptoms of psychotic illness and could be completed by both trained clinicians and patients themselves. The LORS was included in the Massachusetts Guidelines for Schizophrenia as an aid to educate patients about their illness (Commonwealth of Massachusetts, 1999).

Early clinical use of the LORS suggested that numerous patients had insufficient awareness of their psychotic symptoms (Sousa, 2008). This was reflected in higher scores on the LORS when completed by clinicians, as compared to the LORS completed by patients themselves. As an index of these discrepancies, a composite variable was computed by subtracting the LORS–Patient score from the LORS–Clinician score. Conceptually, this discrepancy score reflects the extent to which patients underestimate their psychotic symptoms. The LORS–Discrepancy Scale became the centerpiece of a clinical intervention called the LORS–Enabled Dialogue (Sousa, Corriveau, Lee, Bianco, & Sousa, 2013).

The purpose of the current study was to examine the psychometric properties of the LORS. The primary emphasis was the development of norms and the examination of test-retest reliability. A secondary purpose was to examine the relationship of each of the LORS scales to other published scales purported to measure symptoms of psychotic illness, compliance to medication, and functional assessment.

METHOD

Participants

Ninety patients participated in this study. These patients had diagnoses of schizophrenia, schizoaffective disorder, major depression with psychotic features, or bipolar–I disorder with psychotic features. A best-estimate diagnostic approach was utilized in which information from clinical interviews was supplemented by information from family informants, previous psychiatrists, and medical records. Patients with organic brain disorder or mental retardation were excluded from the study. Patients with intermittent alcohol or substance use were not excluded. Fifty-eight patients were male and 32 were female. Sixty-one patients were outpatients, 28 were inpatients and one began as an outpatient and then became an inpatient. Eighty-two subjects were Caucasian, three were African-American, one was Asian, three were Hispanic and one was in an “other” racial category. Length of illness ranged from less than one year (one person) to 21–25 years (50%). Ages ranged from 19 to 63 years ($M = 43.3$).

Measures

Levels of Recovery Scale (LORS).—The LORS (Sousa, 1998) is a 13-item scale. The items reflected six domains: Need for supervision; Positive

symptoms of illness; Negative symptoms of illness; Work/education activities; Social skills; Self-advocacy. For each item, behavioral descriptions were listed for each of 5 Likert-scale ratings.

The Kemp Compliance Tool (KCT).—The KCT (Kemp, David, & Hayward, 1996) is a method for measuring the extent of adherence to medications. Because patients resided in supervised settings, their compliance with medications was measured by the nurse-clinicians working in these settings. A patient's adherence to medication was assessed through multiple observations before, during, and following medication administration. A 7-point observer-rated compliance scale, with ranges from complete refusal of treatment (higher scores) to active participation with medications, was used to assess compliance. These observations were rated by the primary nurse and by at least two independent mental health professionals who worked with the patient.

The Independent Functional Assessment.—An Independent Functional Assessment based on the work of Morosini, Magliano, Brambilla, Ugolini, and Pioli (2000) included eight items measured on a 7-point Likert scale assessing patient functioning in four areas: socially useful activities; personal and social relationships; self-care; and disturbing and aggressive behaviors. Each item was scored from 1: Total attainment of functioning to 7: No attainment of functioning.

The Positive and Negative Symptom Scale (PANSS).—The PANSS is a clinician-administered measure of the positive and negative symptoms of schizophrenia (Kay, Fiszbein, & Opler, 1987). The PANSS contains 30 items based on a semi-structured clinical interview, and higher scores are indicative of increased symptom severity.

Procedure

Several IRB reviews were completed and approved through the course of this research. All participants provided informed consent. Data was collected from the baseline assessment of 90 patients who participated in a larger treatment efficacy study (Sousa, *et al.*, 2013). Data used to examine test-retest reliability consisted of 40 patients who were assigned to the control condition of the larger study with a one-month interval between the two assessments.

RESULTS

Descriptive statistics for all measures are shown in Table 1. The LORS-Discrepancy Scale data suggests that patients underestimated their symptoms. A paired-comparison *t* test showed that the LORS-Patient scores were significantly lower than the LORS-Clinician scores ($t_{89} = 8.073$, $p < .001$, Cohen's $d = .947$). Table 1 also shows Pearson correlations computed between the first and second administration of all measures. All six scales

TABLE 1
DESCRIPTIVE STATISTICS AND RELIABILITY ESTIMATES FOR ALL INSTRUMENTS

Instrument	N	M	SD	Min.	Max.	Pearson <i>r</i>
Functional Assessment	89	27.91	11.86	4	72	.95
KCT	90	3.12	2.22	1	7	.96
PANSS	89	59.60	16.82	28	102	.95
LORS—Clinician	90	33.39	10.61	13	54	.93
LORS—Patient	90	24.70	7.47	12	47	.98
LORS—Discrepancy	90	8.69	10.21	-17	34	.93

Note.—*N* = 40 for correlation coefficients. All correlations are statistically significant at $p < .001$.

had high test-retest reliability ($r > .90$). The stability of the LORS—Discrepancy scale ($r = .93$) is rather striking in that, as a discrepancy score, it was derived from two scales, each contributing potential error variance.

Pearson correlations computed between all instruments are shown in Table 2. As hypothesized, moderate to high correlations were found between the LORS—Clinician Scale and Functional Assessment (.75), the KCT (.58), and the PANSS (.62). Conversely, the LORS—Patient Scale showed much lower correlations with Functional Assessment (.35), the KCT (.19), and the PANSS (.26). The LORS—Discrepancy Scale showed moderate correlations with Functional Assessment (.52), the KCT (.46), and the PANSS (.46). These results suggest that patients with less awareness of their symptoms showed less appropriate functioning, more problems in complying with their medication regimens, and increased positive and negative symptoms of schizophrenia.

TABLE 2
PEARSON CORRELATION COEFFICIENTS BETWEEN MEASURES

	Functional Assessment	KCT	PANSS	LORS—Clinician	LORS—Patient
Functional Assessment					
KCT	.49†				
PANSS	.57‡	.55‡			
LORS—Clinician	.75‡	.58‡	.62‡		
LORS—Patient	.35‡	.19	.26*	.39‡	
LORS—Discrepancy	.52‡	.46‡	.46‡	.75‡	-.32†

Note.—List-wise deletion resulted in *N* = 88. * $p < .05$. † $p < .01$. ‡ $p < .001$.

DISCUSSION

Collectively, the results of this study illustrate that the LORS—Clinician Scale, the LORS—Patient Scale, and the LORS—Discrepancy Scale all

showed high test-retest reliability. The LORS–Patient Scale showed the highest concurrent validity with published scales. The LORS was not intended to overlap perfectly with Functional Assessment, the KCT, or the PANSS, and the correlation coefficients found in this study were in the desired range. One advantage of the LORS is its efficiency. The LORS took about half the time to complete than the PANSS. The most important and notable advantage of the LORS is the extent to which it can be completed by both trained clinicians and by patients themselves.

As expected, the LORS–Patient Scale scores were the most weakly correlated to other measures. The most obvious explanation is that many patients were not sufficiently aware of their psychotic symptoms: LORS–Patient Scale scores were significantly lower than LORS–Clinician Scale scores. The data presented for the LORS–Discrepancy Scale further demonstrated the extent to which patients underestimated their pathology and symptoms. This discrepancy scale may assist clinicians in identifying patients with poor self-awareness of psychotic symptoms. A dialogue or discussion of these discrepancies may be the focus of effective low-cost treatment (Sousa, *et al.*, 2013). While the LORS–Clinical Scale may be the most valid in identifying psychotic illness, the LORS–Discrepancy Scale may be the most useful in identifying patients whose lack of awareness of their psychotic symptoms is not readily identified in a typical therapy session. These patients may potentially benefit the most from treatment designed to discuss these discrepancies.

One potential limitation of this study is the multiple diagnostic categories. While all participants were selected based on the presence of psychotic symptoms, more research is needed to understand the LORS–Discrepancy Scale scores in different diagnostic categories. Similarly, all participants in this study were receiving medication. Future research should include patients who are not prescribed medication. Lastly, more research is needed to understand which diagnostic group might benefit most from a “discrepancy-driven” therapeutic dialogue.

REFERENCES

- BECK, E., CAVELTI, M., KVRGIC, S., KLEIM, B., & VAUTH, R. (2011) Are we addressing the “right stuff” to enhance adherence in schizophrenia? Understanding the role of insight and attitudes towards medication. *Schizophrenia Research*, 132, 42-49. DOI: 10.1016/j.schres.2011.07.019
- CHAKRABORTY, K., & BASU, D. (2010) Insight in schizophrenia: a comprehensive update. *German Journal of Psychiatry*, 13, 17-30.
- COMMONWEALTH OF MASSACHUSETTS. (1999) *Clinical practice guidelines for the treatment of schizophrenia in adults*. Retrieved from <http://www.mass.gov/eohhs/docs/dmh/publications/schizophrenia-guidelines-full.pdf>.
- CRAMER, J., & ROSENHECK, R. (1998) Compliance with medication regimes for mental and physical disorders. *Psychiatric Services*, 49, 196-201.

- GAEBEL, W., SCHREINER, A., BERGMANS, P., DE ARCE, R., ROUILLON, F., CORDES, J., & SMERALDI, E. (2010) Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable quetiapine: results of a long-term, open-label, randomized clinical trial. *Neuropsychopharmacology*, 35(12), 2367-2377. DOI: 10.1038/npp.2010.111
- KAY, S., FISZBEIN, A., & OPLER, L. A. (1987) The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13, 261-276.
- KEMP, R., DAVID, A., & HAYWARD, P. (1996) Compliance therapy: an intervention targeting insight and treatment adherence in psychotic patients. *Behavioural and Cognitive Psychotherapy*, 24, 331-350.
- LACRO, J. P., DUNN, L. B., DOLDER, C. R., LECKBAND, S. G., & JESTE, D. V. (2002) Prevalence of and risk factors for medication nonadherence in patients with schizophrenia: a comprehensive review of recent literature. *Journal of Clinical Psychiatry*, 63, 892-909.
- LINCOLN, T. M., LÜLLMANN, E., & RIEF, W. (2007) Correlates and long-term consequences of poor insight in patients with schizophrenia: a systematic review. *Schizophrenia Bulletin*, 33, 1324-1342. DOI: 10.1093/schbul/sbm002
- LINDEN, M., & GODEMANN, F. (2007) The differentiation between 'lack of insight' and 'dysfunctional health beliefs' in schizophrenia. *Psychopathology*, 40, 236-241. DOI: 10.1159/000101366
- MASAND, P., ROCA, M., TURNER, M., & KANE, J. M. (2009) Partial adherence to antipsychotic medication impacts the course of illness in patients with schizophrenia: a review. *Journal of Clinical Psychiatry*, 11, 147-154. DOI: 10.4088/PCC.08r00612
- MOROSINI, P. L., MAGLIANO, L., BRAMBILLA, L., UGOLINI, S., & PIOLI, R. (2000) Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatrica Scandinavica*, 101, 323-329.
- NOVICK, D., HARO, J., SUAREZ, D., PEREZ, V., DITTMANN, R. W., & HADDAD, P. M. (2010) Predictors and clinical consequences of non-adherence with antipsychotic medication in the outpatient treatment of schizophrenia. *Psychiatry Research*, 176, 109-113. DOI: 10.1016/j.psychres.2009.05.004
- SOUSA, S. (1998) Levels of recovery from psychotic disorders chart. *Journal of Psychosocial Nursing*, 37, 19-23.
- SOUSA, S. (2008) The LORS-enabled dialogue: improving insight in psychotic illnesses. *Journal of Psychosocial Nursing and Mental Health Services*, 46, 44-51.
- SOUSA, S. A., CORRIVEAU, D., LEE, A. F., BIANCO, L. G., & SOUSA, G. M. (2013) The LORS-enabled dialogue: a collaborative intervention to promote recovery from psychotic disorders. *Psychiatric Services*, 64, 58-64.

Accepted August 22, 2013.

FACTORIAL STRUCTURE AND CONVERGENT AND DISCRIMINANT VALIDITY OF THE E (EMPATHY) SCALE¹

ULRICH S. TRAN

*Department of Basic Psychological Research and Research Methods
School of Psychology, University of Vienna*

ANTON-RUPERT LAIREITER

*Department of Psychology
University of Salzburg*

DAVID P. SCHMITT

*Department of Psychology
Bradley University, Peoria, Illinois*

CHRISTINE NEUNER

*Department of Psychology
University of Salzburg*

MAX LEIBETSEDER

*Department of Social Psychiatry,
State Government of Salzburg*

SARA LEYLA SZENTE-VORACEK

Vienna

MARTIN VORACEK

*Department of Basic Psychological Research and Research Methods
School of Psychology, University of Vienna
Georg Elias Müller Department of Psychology
Georg August University of Göttingen*

Summary.—The Empathy (E) scale has been proposed as a theoretically and psychometrically more satisfying alternative to existing self-report measures of empathy. Its four scales (facets) cover both components (cognitive vs. emotional) and both reality statuses (fictitious vs. real-life) of empathy in pairwise combinations. Confirmatory factor analyses of the E-scale in an Austrian community sample ($N = 794$) suggested that one prior assumption, namely the mutual orthogonality of these facets, may partly need revision; particularly, the E-scale facets seemed to reflect more strongly differences in the reality statuses than in the components of empathy. Utilizing numerous informative psychological traits, the scale's convergent and discriminant validity were examined. E-scale scores were consistently predicted by sex-related and relationship-related constructs and measures of anti-social attitudes and behavior. Among the Big Five personality dimensions, open-

¹Address correspondence to Martin Voracek, D.Sc., D.M.Sc., Ph.D., Department of Basic Psychological Research and Research Methods, School of Psychology, University of Vienna, Liebiggasse 5, Room 03-46, A-1010 Vienna, Austria, or Georg Elias Müller Department of Psychology, Georg August University of Göttingen, Goßlerstraße 14, D-37073 Göttingen, Germany, or e-mail (martin.voracek@univie.ac.at). Thanks are extended to Therese Feichter and Lisa Mariella Loibl for assistance with this research, and to Clara Blum (née Ertl), Ulrike Pum, and Barbara Rupp for help with data collection.

ness emerged as a major positive correlate of empathy. Sex and age were demographic correlates of E-scale scores (higher in women and the younger). Findings were discussed with regards to the definition and measurement of empathy.

Empathy is an important but diverse psychological construct which is associated with prosocial and altruistic behaviors (Eisenberg & Miller, 1987). Lack of empathy signifies one clinical symptom of narcissistic personality disorder, deficits in empathy are characteristic for autism-spectrum disorders, and low empathy is also linked to violent and antisocial behavior (see DSM-IV-TR; APA, 2000; Geer, Estupinan, & Manguno-Mire, 2000). Whilst the study of empathy and empathic behavior seems relevant for various psychological fields, the respective assessment via self-report measures appears somewhat compromised and leaves both researchers and theorists with some unresolved problems regarding the structure and necessary components of empathy. In what follows, we provide a brief account on the measurement of empathy, and the remainder of this article will focus on a newer measure, namely the E-scale (Leibetseder, Laireiter, Riepler, & Köller, 2001; Leibetseder, Laireiter, & Köller, 2007).

Of the considerable number of self-report measures of empathy, only a few have been deemed psychometrically reliable and are more widely used (Jolliffe & Farrington, 2004). Historically, different scientific traditions and approaches influenced the study of empathy (for an overview, see Baron-Cohen & Wheelwright, 2004), whence widely-used measures in this area, many of which by now are about 30 years old, are based on seemingly different and competing models and conceptualizations. Whilst some measures focus on emotional empathy (Questionnaire Measure of Emotional Empathy; Mehrabian & Epstein, 1972; Impulsiveness-Venturesomeness-Empathy Scale; Eysenck & Eysenck, 1978; Index of Empathy for Children and Adolescents; Bryant, 1982; Toronto Empathy Questionnaire; Spreng, McKinnon, Mar, & Levine, 2009), others focus on cognitive empathy (Hogan Empathy Scale; Hogan, 1969). Emotional (or affective) empathy broadly encompasses the emotion-contagion hypothesis (Mehrabian & Epstein, 1972), defining empathy by matching or in some other way appropriate emotional responses of observers to observed emotional states of others (Baron-Cohen & Wheelwright, 2004). Cognitive empathy focuses on the understanding of others perspectives, as, for example, described in Hogan's (1969) perspective-taking hypothesis, but may also be likened to the theory of mind concept in general (Premack & Woodruff, 1978).

Davis (1983) and Baron-Cohen and Wheelwright (2004) maintained that empathy is a multidimensional construct which incorporates both emotional and cognitive components that may, however, not be easily separated but rather are indispensable parts of the same construct. Widely-used

measures devised by these researchers (Interpersonal Reactivity Index; Davis, 1983; Empathy Quotient; Baron-Cohen & Wheelwright, 2004) allow the assessment of both emotional and cognitive empathy, avoiding a possibly too-narrow trait coverage.

Empirically, the Interpersonal Reactivity Index (cognitive empathy: perspective taking and fantasy; emotional empathy: empathic concern and personal distress) has been found to be dominated by a general factor which, in turn, is conceptually identical to empathic concern (Cliffordson, 2001, 2002) or grouped together in one superfactor (Pulos, Elison, & Lennon, 2004). The cognitive and emotional scales and factors of the Empathy Quotient are also substantially interrelated (correlating about .50 or higher; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004; Muncer & Ling, 2006). Thus, emotional and cognitive empathy, even though postulated to be conceptually different, seem to be structurally closely related.

Regarding empathy deficits in Asperger syndrome, emotional and cognitive empathy appear to be functionally differentiated (Rogers, Dziobek, Hassenstab, Wolf, & Convit, 2007). Yet, in other relevant areas of empathy research, e.g., with regard to offending behavior (Jolliffe & Farrington, 2004), emotional and cognitive empathy could not be convincingly disentangled, for which problems with the content and construct validity of widely-used measures were mainly held responsible.

Generally, existing measures of empathy have been criticized repeatedly with regard to their validity. For example, the Hogan Empathy Scale was constructed to measure how much a respondent resembles a "highly empathic man" (Hogan, 1969, p. 308). Whilst this criterion must be considered likely outdated (Jolliffe & Farrington, 2004) and questionable, the interpretation of test scores (or rather, similarity ratings) may also lead to tautologies that are deprived of psychological meaning (Kline, 1992). Further content validity problems of the Hogan Empathy Scale have been highlighted by Holz-Ebeling and Steinmetz (1995), Baron-Cohen and Wheelwright (2004), and Jolliffe and Farrington (2004).

Similar criticism concerning content validity was also brought forward against the Questionnaire Measure of Emotional Empathy and the Interpersonal Reactivity Index; namely, that their specific operational definitions of empathy differ from their purported theoretical definitions (Holz-Ebeling & Steinmetz, 1995) and that they are confounded measures of empathy (Baron-Cohen & Wheelwright, 2004). Most recently, the Questionnaire of Cognitive and Affective Empathy (Reniers, Corcoran, Drake, Shryane, & Völlm, 2011) has been introduced. However, this measure merely builds on a compilation of the item pool of existing measures (Empathy Quotient, Hogan Empathy Scale, Interpersonal Reactivity Index, and Impulsiveness-Venturesomeness-Empathy Scale), which were

subjected to expert ratings of item content appropriateness. About one-half (15 items) of the final, 31-item form of the Questionnaire of Cognitive and Affective Empathy comprises items from the Empathy Quotient, a measure that is so strongly correlated ($r < .70$; Nettle, 2007) with agreeableness as to suggest that it is essentially equivalent with this dimension from the Big Five factor model of personality.

As the relationship between emotional and cognitive empathy is still debated, a further distinction has so far, for the most part, been neglected: empathy in fictitious (e.g., movies and texts) and real-life situations. Due to mere idealization, fictitious situations may elicit highly empathic reactions, which may not be as easily aroused in real-life situations. Shapiro and Rucker (2004) coined the term "Don Quixote effect" as a framework to foster a form of emotional idealism elicited by having real medical practitioners watch medical TV series and movies to promote empathy and altruism in their everyday practice. In terms of assessment, these two areas (fictitious and real-life) should be differentiated systematically, since test scores may otherwise misestimate real-life empathy. Of the available measures, only the Interpersonal Reactivity Index's Fantasy scale allows the assessment of idealizing empathic tendencies. However, it also contains construct-irrelevant items and contents (e.g., "I daydream and fantasize, with some regularity, about things that might happen to me") and is only assumed to represent cognitive empathy in the framework of the Interpersonal Reactivity Index. Hence, for a more thorough investigation of the structure of empathy (also accounting for possible differences between real-life and fictitious situations, and between cognitive and emotional empathy), existing measures clearly do not suffice.

The E-scale

Leibetseder and his associates (Leibetseder, *et al.*, 2001, 2007; full item texts are provided in the latter) introduced the E-scale as a new self-report measure of empathy. Starting from existing questionnaires (Hogan Empathy Scale, Questionnaire Measure of Emotional Empathy, Interpersonal Reactivity Index, as well as a scale of Stotland, Kenneth, Mathews, Sherman, Hansson, & Richardson, 1978), the E-scale comprises 25 items, scored on five-point scales, which capture emotional as well as cognitive components of empathy and also distinguish between empathic reactions in fictitious and real-life situations. Leibetseder, *et al.* (2007) fitted a nested-factors model to the E-scale, wherein all items load on a general factor of empathy and, additionally, group into four specific factors according to the factorial combinations of component (cognitive vs. emotional) and reality status (fictitious vs. real-life). Accordingly, the E-scale was proposed as a measure of general empathy (when its total score is used) as

well as of four facets of empathy (when its four scale scores are used), namely Cognitive Sensitivity (S1; cognitive empathy, fictitious situation), Emotional Sensitivity (S2; emotional empathy, fictitious situation), Emotional Concern (S3; emotional empathy, real-life situation), and Cognitive Concern (S4; cognitive empathy, real-life situation).

Convergent validity of the E-scale has been demonstrated through positive and mostly strong correlations of its scales with IVE scores and measures of social intelligence, prosocial orientation, responsiveness in intimate relationships, and anxiety, and through negative correlations with self-esteem (Leibetseder, *et al.*, 2007). Discriminant validity of the E-scale has been shown considering various measures of perceived regard from others and partner satisfaction. However, these validity studies were based on rather small samples (*Ns* ranging from 60 to 160), and associations with further potentially interesting constructs, as well as sex differences or age effects, have up to now not been investigated.

Present Research

This study set out to independently replicate the factorial structure of the E-scale in a large community sample. Given the above-mentioned findings of equivocal psychometric structure of some other multidimensional empathy scales, this intention seems paramount. Validation of the structure of the E-scale in yet another community sample (the main validation sample for the E-scale was also a large community sample; Leibetseder, *et al.*, 2007) may further strengthen the standing of this measure, since most other validation efforts used undergraduate samples (Jolliffe & Farrington, 2004).

Another point of consideration pertains to the methods used in the structural analyses of both the E-scale and other measures of empathy. Given the ordinal response format of the items of the E-scale (as well as of other measures), methods for continuously scored items, as implicitly assumed by common factor-analytical approaches, do not seem appropriate. Instead, factor-analytical methods for ordered categorical items, based on the polychoric correlation matrix of items, appear to be more appropriate and also may guard against spurious results. Thus, this study also serves to test previous results on the structure of the E-scale with more appropriate statistical methodology.

Furthermore, associations of the E-Scale with respondent sex, age, and some constructs commonly reported to be associated with empathy were analyzed for the first time. Because women on average report higher levels of empathy (Eisenberg & Lennon, 1983), respondent sex was a primary variable of interest in this context. It was predicted that women score higher on the E-scale than men. As for examples of the further variables utilized for testing the construct validity of the E-scale, there is, for instance, no good

evidence for whether affiliation to prescriptive (stereotyped) sex roles also yields comparable group differences in empathy. It was predicted that less prescriptive sex-role orientations would be associated with higher E-scale scores. Positive regard of others (within or outside of intimate relationships), but not relationship quality or sociosexuality, was predicted to be positively associated with empathy. Because violent and antisocial behaviors clearly constitute antipodes to empathic reactions (Miller & Eisenberg, 1988; Cohen & Strayer, 1996; Jolliffe & Farrington, 2004), it was predicted that these would correlate negatively with the E-scale, which was also assumed for narcissism and social dominance orientation (Watson, Grisham, Trotter, & Biderman, 1984; Pratto, Sidanius, Stallworth, & Malle, 1994). Furthermore, in a similar vein based on prior empirical evidence, it was predicted that higher scores on the E-scale would be associated with the tendency to respond in a less socially desirable manner (Steins, 1998) and with higher levels of neuroticism (Eysenck & Eysenck, 1978) and agreeableness (Graziano & Eisenberg, 1997). In exploring these correlates of empathy, it was documented that even though the E-scale lives up to most of its expected correlates, some results lead to expectations and concerns that need further research.

METHOD

Participants and Procedure

Respondents were enrolled within a larger survey, namely the ISDP-2, which was a follow-up study of the ISDP (International Sexuality Description Project; for project background and details, including survey translation procedures, see Schmitt, Alcalay, Allensworth, Allik, Ault, Austers, *et al.*, 2003, 2004; Schmitt, Allik, McCrae, & Benet-Martinez, 2007; Schmitt, Realo, Voracek, & Allik, 2008) that was conducted in more than 60 countries around the world. The ISDP-2 investigated numerous relationships between personality dimensions and sexuality and, among other goals, explored sex differences in mating-related behavior and preferences as well. The survey form administered across nations and cultures comprised 22 different psychological scales and measures, with the E-scale added to these at the Austrian study sites. In total, 822 residents from the cities of Salzburg and Vienna participated. A total of 794 participants provided sufficient data on the E-scale to be included in the analysis (Salzburg: $N = 410$; Vienna: $N = 384$), of which 50.3% (399 respondents) were women (age $M = 31.7$ years, $SD = 11.0$, range: 18–67 years). Participants had on average 14.7 ($SD = 3.2$) years of education.

Structural Analysis

We tested the confirmatory factor-analytic models of Leibetseder, *et al.* (2007; Models 1, 2, and 3) and also four further models (henceforth,

Models 2a, 3a, 4, and 5). Model 1 comprised a general (g) factor of empathy; Model 2 was a group-factor model (i.e., no g-factor), wherein each scale was represented by a single factor and all factors were allowed to correlate; Model 3 was a nested-factors model with a g-factor and specific factors S1 to S4, wherein the specific factors were assumed to be orthogonal to each other and to the g-factor (see Table 3 for the assignment of items to scales S1 to S4; see Brunner, Nagy, & Wilhelms, 2012, for background and discussion of these confirmatory factor-analytic models).

Apparently, Model 2 might be more parsimoniously represented by a hierarchical factor model, wherein S1 to S4 form a first stratum and empathy is a superfactor on top of this stratum (Model 2a). Notably, the four scales were assumed to measure distinct facets (i.e., factorial combinations of reality status and component) of empathy; hence, Model 2a may also assess this assumption more directly than Models 2 or 3. Moreover, the four facets were hypothesized to incorporate aspects of empathy that are identical across pairs of scales (e.g., S1 and S4 both capture cognitive empathy, whilst at the same time S1 and S2 both pertain to fictitious situations). In light of this, Model 3 seems unsuitable from a theoretical point of view, as mutual orthogonality of the scales is implied therein. Allowing for non-zero correlations among the specific factors (Model 3a) should match theoretical assumptions better. However, Model 3a did not properly converge with our data (see Results). Hence, Models 4 and 5 were also tested: Combining S1 with S4 (Cognitive Empathy) and S2 with S3 (Emotional Empathy), a two-factor model was fitted to the data in order to test the extent to which the E-scale represents these two components of empathy (Model 4). Model 5 combined S1 with S2 (fictitious situations, reflecting sensitivity) and S3 with S4 (real-life situations, reflecting concern). Comparable fits of Model 4 and 5 would indicate that both reality status and components of empathy are equally well represented and assessed by the E-scale.

In contrast to Leibetseder, *et al.* (2007), who opted to omit Items 12 and 16, these were included in the present structural analyses in order to test whether their exclusion is warranted. Item 12 was assigned to S2 and Item 16 to S4.

Mplus 6.1 (Muthén & Muthén, 2008) was used for all confirmatory factor analyses, utilizing the WLSMV estimator, which assumes ordered categorical items and provides robust standard errors and model-fit tests. Model fit was evaluated using a number of widely applied fit indices, with cut-off values for good and adequate model fit as recommended by Hu and Bentler (1999): CFI (comparative fit index), TLI (Tucker-Lewis index; good fit indicated by CFI or TLI values $\geq .95$, and adequate fit by $\geq .90$), and RMSEA (root mean square error of approximation, with good or adequate fit indicated by values of $< .06$ and $< .08$, respectively).

Construct Validation Instruments and Predictions

Fifteen of the 22 scales contained in the ISDP-2 survey were deemed as potentially informative in this context and were thus included as suited for the construct validity analyses of the E-scale. For a number of variables, their relationships with empathy have already been established on empirical grounds (see above) or may be assumed according to theoretical considerations. Convergent constructs (for which positive or negative correlations with empathy were predicted) and discriminant constructs (for which null correlations with empathy were predicted) are listed below.

Positive correlations were predicted for the agreeableness and neuroticism dimensions of the Big Five (as assessed with the BFI; John & Srivastava, 1990) for a perspective on sex roles emphasizing equality and opposing prescriptive beliefs about the behaviors and roles of men and women (SRI; Kalin & Tilby, 1978; sample item: "Women should be allowed the same sexual freedom as men"), for a more feminine gender orientation, and for partnership-oriented emotional investment (both taken from the Sexy Seven [S7] inventory of Schmitt & Buss, 2000).

Negative correlations were predicted for sexual violence (ASBI; Mosher, 1998), socially desirable responding (the impression management subscale of the BIDR; Paulhus, 1991), Machiavellianism (MACH-IV; Christie & Geis, 1970), narcissism (NPI; Raskin & Hall, 1979), rape-supportive attitudes (RSA; Lottes, 1988; reverse-scored for consistency of analysis), social dominance orientation (SDO; Pratto, *et al.*, 1994), self-esteem (SES; Rosenberg, 1965), psychopathy (SRP-III; Williams, Nathanson, & Paulhus, 2003), and nonsexual violence (VAI; Dobash, Dobash, Cavanagh, & Lewis, 2000). Null correlations were predicted for the conscientiousness, extraversion, and openness dimensions of the Big Five: self-deceptive overconfidence in one's own judgments and rationality (self-deceptive positivity subscale of the BIDR), subjective well-being (SWB; Bradburn, 1969), and a number of measures pertaining to relationship and sexuality, namely relationship quality (PRQC; Fletcher, Simpson, & Thomas, 2000), sociosexuality (SOI; Simpson & Gangestad, 1991), and the erotophilic disposition, relationship exclusivity, sexual attractiveness, sexual restraint, and sexual orientation scales from the S7 inventory (Schmitt & Buss, 2000).

Construct Validation Analysis

Correlational analyses (based on Pearson's r) were carried out to explore the convergent and discriminant validity facets of the E-scale (for total and subscale scores). Since noticeable sex differences in these scores were expected, in order to guard against spurious findings all correlations were computed separately for the sexes. Standardized mean score differ-

ences (Cohen's d values) between the sexes and significance tests of these group effects (unpaired t tests) are provided. Significance was set to $p < .05$ (two-tailed) for these tests, and $p < .01$ (one-tailed for predicted positive or negative correlations, two-tailed for predicted null correlations) for correlational analyses to guard against inflation of Type I error rates.

RESULTS

Dimensionality of the E-scale

Internal consistency (Cronbach's α) was .92 for the total scale, and .88, .87, .76, and .66 for the scales S1 to S4. Table 1 shows the results of the fit of the different factor structures. Due to their redundancy in content, error terms of item pairs Item 2 ("If I see a good movie, I can easily feel like the principal actor") and Item 7 ("In a good movie I can easily put myself in the principal actor's place"), and of item pairs Item 5 ("I can easily relive the feelings of characters in a novel") and Item 9 ("I can really deeply understand the feelings of the characters in a novel"; all belonging to S2) were allowed to correlate in all models in order to achieve proper convergence. As compared with the results of Leibetseder, *et al.* (2007), Models 1, 2, and 3 yielded only a mediocre fit to the data (see Table 1). In contrast to prior findings, the group-factor model (Model 2) had a better fit than the nested-factors model (Model 3). However, within the nested-factors model, all items loaded significantly ($ps < .001$) on the g-factor of empathy

TABLE 1
MODEL-FIT STATISTICS OF CONFIRMATORY FACTOR ANALYSIS MODELS

Model	χ^2	df	CFI	TLI	RMSEA [90% CI]
1: g-factor model	3120.70	273	.885	.874	.115 (.111, .118)
2: Four correlated-factors model	1490.23	267	.951	.945	.076 (.072, .080)
2a: Hierarchical model	2158.77	269	.924	.915	.094 (.090, .098)
3: Nested-factors model, orthogonal specific factors	1968.86	248	.931	.916	.093 (.090, .097)
3a: Nested-factors model, correlated specific factors	No convergence				
4: Two-factor model (Cognitive and Emotional Empathy)	3042.59	272	.888	.877	.113 (.110, .117)
5: Two-factor model (Sensitivity and Concern)	1673.25	272	.944	.938	.081 (.077, .084)

Note.—All $ps < .001$ for χ^2 values. The best-fitting models are printed in boldface.

TABLE 2
FACTOR INTERCORRELATIONS IN MODEL 2

Factor	S1	S2	S3
S2	.90		
S3	.65	.60	
S4	.74	.73	.92

Note.—All $ps < .001$.

(mean loading = .57, Range = .25 to .81), suggesting that the use of a total score is warranted with regards to the E-scale.

In the best-fitting of these models, namely Model 2, factor intercorrelations were generally high (Table 2) and their pattern matched previous findings. However, factors S1 and S2 (Sensitivity), and S3 and S4 (Concern) intercorrelated highly and were nearly indistinguishable within pairs, suggesting that the scales of the E-scale distinguished more strongly between reality statuses than between components of empathy. The error terms of Items 2 and 7, and of Items 5 and 9, correlated with .63 and .50 ($ps < .001$), respectively. The factor loadings of Model 2 are displayed in Table 3. Notably, Items 12 and 16 had acceptable loadings and hence were kept in their respective scales. The lowest loading was observed for Item 20 (.33; S3), and most items of S4 had only medium-sized loadings (range = .37 to .69). The relatively low discriminating power of these items may thus explain the modest reliability of S4 (see above).

As the hierarchical model (Model 2a) fitted worse than the group-factor model (Model 2; Table 1), intercorrelations among the four scales were obviously not fully captured by a single superfactor. Loadings of the scales S1 to S4 on higher-order Empathy were .95, .91, .74, and .88 (all $ps < .001$). Model 3a, allowing for correlations among the four scales in the nested-factor model, failed to converge. Model 5 fitted the data considerably better than Model 4 (see Table 1) and also came close in fit to Model 2. This again suggests that the E-scale more clearly differentiates sensitivity from concern than cognitive from emotional empathy. Accordingly, Cognitive Empathy (S1 and S4) and Emotional Empathy (S2 and S3) correlated with .94 ($p < .001$) in Model 4, whilst Sensitivity (S1 and S2) and Concern (S3 and S4) correlated with .70 ($p < .001$) in Model 5. Factor loadings in Model 5 are provided in Table 3. Notably, combining raw scores of S1 with S2 (Sensitivity), and of S3 with S4 (Concern), yielded scales with internal consistencies of .92 and .83. Hence, the use of a total score and of these two scales may be recommended for the E-scale and will be followed throughout the subsequent analyses.

TABLE 3
ITEM FACTOR LOADINGS IN MODEL 2 AND MODEL 5

Item	Model 2				Model 5	
	Cognitive Sensitivity (S1)	Emotional Sensitivity (S2)	Emotional Concern (S3)	Cognitive Concern (S4)	Sensitivity (S1 and S2)	Concern (S3 and S4)
21	.86				.84	
19	.86				.84	
24	.82				.81	
17	.80				.78	
14	.75				.74	
7		.85			.82	
9		.84			.81	
12		.84			.81	
5		.82			.79	
2		.75			.72	
11		.49			.48	
1		.41			.40	
15			.78			.75
18			.75			.72
23			.67			.63
10			.59			.57
22			.56			.54
25			.53			.51
20			.33			.32
6				.69		.71
16				.62		.63
8				.59		.60
13				.49		.50
3				.45		.45
4				.37		.38

Note.—Within subscales (S1 to S4), items are listed in descending order with reference to their loadings. All $ps < .001$.

Sex and Age Effects

Women's self-reported levels of empathy markedly exceeded those of men's: Cohen's d values were -0.72 , -0.52 , and -0.78 (all $ps < .001$) for the total score and the scale scores of Sensitivity and Concern; i.e., women exceeded men's self-reported empathy even more strongly with respect to real-life situations. Respondent age was consistently negatively associated with E-Scale scores (in above order): $r = -.20$, $-.22$, and $-.14$ (all

$ps < .001$), meaning that sensitivity decreased slightly stronger than concern with age.

Convergent and Discriminant Validity

There were sex differences (mostly of small to medium size) for the majority of the assumed convergent and discriminant constructs of the E-scale (Table 4). Predicted positive correlations were mostly small to medium and were occasionally differentiated for sex or reality status. Whilst less prescriptive sex-role beliefs and emotional investment were correlates of empathy across the sexes and all three scores of the E-scale (total score, Sensitivity, Concern; being slightly stronger among women than men), agreeableness and neuroticism were convergent correlates only among men. Among women, agreeableness showed a small association only with Concern. Feminine gender orientation also had a slight influence on men's scores, but only for the Sensitivity component.

Predicted negative correlations were also mostly small to medium and were generally differentiated for sex and reality status. Only social dominance orientation was consistently negatively associated with all three scores across both men and women. Machiavellianism showed only negative associations with Concern (slightly stronger for men than women). Sexual violence was a weak negative correlate of Concern only among men, while rape-supportive attitude was a negative correlate of Concern and Sensitivity only among women. Psychopathy (SRP-III total score, without the items of the Cold Affect scale; see below) was a medium-sized negative correlate of Concern among men, but a weak *positive* correlate of Sensitivity among women. Yet, as was apparent from the scale scores of the SRP-III, this positive association was only due to the ratings on the measures's impulsive thrill-seeking scale. Interpersonal manipulation was the strongest negative correlate of Concern (and also of Sensitivity among men) for both men and women. Since the cold affect scale of the SRP-III had unacceptable measurement properties (Cronbach's $\alpha = .05$) in this sample, the items of this scale were excluded from analysis.

Unexpectedly, nonsexual violence was also found to be weakly *positively* associated with Sensitivity and Concern among women. Further analysis revealed that this was due to conspicuous ratings on merely two items (the VAI assesses 20 violent acts of respondent against current partner, e.g., "Pushed, grabbed or shoved partner," and is scored from 0 = *never*, 1 = *once*, 2 = *twice*, 3 = *3 to 5 times*, 4 = *6 to 10 times*, 5 = *11 or more times*): "Shout at partner" and "Swear at partner," which were the only items that yielded means > 1 among both men and women (overall $M = 1.92$ and 1.31 , respectively), whilst total VAI scores were very low (men: $M = 0.25$, $SD = 0.40$, women: $M = 0.30$, $SD = 0.36$). Positive associations

TABLE 4
ASSOCIATIONS OF THE E-SCALE WITH CONVERGENT AND DISCRIMINANT CONSTRUCTS

Measure	α	Cohen's d	Total Score		Sensitivity		Concern	
			Men	Women	Men	Women	Men	Women
Predicted positive correlations								
Agreeableness (BFI)	.69	−0.16*	.18‡	.07	.12‡	.01	.21‡	.13‡
Neuroticism (BFI)	.82	−0.52‡	.23‡	.09	.17‡	.05	.24‡	.11
Less prescriptive sex-role beliefs (SRI)	.67	−0.55‡	.18‡	.27‡	.18‡	.26‡	.13‡	.23‡
Emotional investment (S7)	.84	−0.28‡	.21‡	.29‡	.15‡	.24‡	.23‡	.29‡
Feminine gender orientation (S7)	.83	−2.91‡	.12‡	.04	.15‡	.02	.07	.06
Predicted negative correlations								
Sexual violence (ASBI)	.73	0.43‡	−.11	.01	−.07	.08	−.14‡	−.08
Impression management (BIDR)	.73	−0.24‡	.02	−.13‡	−.03	−.17‡	.08	−.04
Machiavellianism (MACH-IV)	.78	0.34‡	−.19‡	−.12‡	−.09	−.05	−.26‡	−.18‡
Narcissism (NPI)	.85	0.35‡	−.05	.10	−.00	.12	−.10	.06
Self-sufficiency	.45	0.43‡	−.13‡	−.03	−.10	−.01	−.13‡	−.05
Rape-supportive attitude (RSA) ^a	.76	0.49‡	−.12‡	−.17‡	−.11	−.13‡	−.09	−.18‡
Social dominance orientation (SDO)	.65	0.34‡	−.26‡	−.23‡	−.20‡	−.17‡	−.27‡	−.25‡
Self-esteem (SES)	.96	0.01	−.09	.06	−.03	.05	−.13‡	.06
Psychopathy (SRP-III)	.82	0.56‡	−.13‡	.05	−.04	.14‡	−.22‡	−.08
Antisocial behaviour	.69	0.41‡	−.09	.04	−.02	.11	−.15‡	−.06
(continued on next page)								

(continued on next page)

Note.—Analysis $N = 373\text{--}395$ (men) and $373\text{--}399$ (women), except for VAI (309 men, 310 women) and PRQC (280 men, 290 women). Positive d values indicate higher mean scores for men. ^aReverse-scored. Correlations with $p < .01$ are highlighted with shaded fields. $+p < .10$, $*p < .05$, $\dagger p < .01$, $\ddagger p < .001$, two-tailed for mean differences, one-tailed for predicted negative and positive correlations (with the exception of tests involving scales of the SRP-III and VAI; see text for details).

TABLE 4 (CONT'D)
ASSOCIATIONS OF THE E-SCALE WITH CONVERGENT AND DISCRIMINANT CONSTRUCTS

Measure	α	Cohen's d	Total Score		Sensitivity		Concern	
			Men	Women	Men	Women	Men	Women
Impulsive thrill-seeking	.77	0.54‡	-.04	.14†	.00	.19‡	-.08	.03
Interpersonal manipulation	.61	0.44‡	-.23‡	-.08	-.13†	.01	-.30‡	-.18‡
Nonsexual violence (VAI)	.72	-0.14 ⁺	.01	.19‡	.03	.18‡	-.03	.17‡
Predicted null correlations								
Conscientiousness (BFI)	.79	-0.02	-.09	-.10	-.12	-.13†	-.04	-.02
Extraversion (BFI)	.84	-0.15*	-.15†	.07	-.13	.03	-.15†	.10
Openness (BFI)	.80	-0.09	.31‡	.27‡	.32‡	.25‡	.21‡	.23‡
Self-deceptive positivity (BIDR)	.70	0.41‡	-.25‡	-.05	-.23‡	-.05	-.20‡	-.03
Relationship quality (PRQC)	.88	0.07	-.01	.13	-.02	.09	.00	.15
Well-being (SWB)	.70	0.13 ⁺	-.07	-.02	-.05	-.05	-.07	.02
Sociosexuality (SOI)	.61	0.34‡	-.04	.04	-.01	.10	-.06	-.04
Erotophilic disposition (S7)	.86	0.51‡	.00	.12	.02	.14†	-.03	.06
Relationship exclusivity (S7)	.79	-0.33‡	.14†	.03	.11	-.02	.15†	.08
Sexual attractiveness (S7)	.90	-0.31‡	.10	.17‡	.09	.14†	.07	.18‡
Sexual restraint (S7)	.69	-0.10	-.02	-.07	-.02	-.05	-.03	-.09
Sexual orientation (S7)	.67	-0.05	.07	.04	.08	.06	.04	-.00

Note.—Analysis $N = 373$ – 395 (men) and 373 – 399 (women), except for VAI (309 men, 310 women) and PRQC (280 men, 290 women). Positive d values indicate higher mean scores for men. ^aReverse-scored. Correlations with $p < .01$ are highlighted with shaded fields. ⁺ $p < .10$, ^{*} $p < .05$, [†] $p < .01$, [‡] $p < .001$, two-tailed for mean differences, one-tailed for predicted negative and positive correlations (with the exception of tests involving scales of the SRP–III and VAI; see text for details).

among women could be traced to these two items exclusively, whereas none further to the remainder of items portraying other acts of violence (details omitted for brevity).

Impression management appeared to be negatively associated with Sensitivity (only among women). Self-esteem was a weak negative correlate of Concern (only among men). Of note, narcissism (NPI total scores) appeared to be unrelated to empathy, as measured by the E-scale. Only for one of the eight NPI scales, namely self-sufficiency, could a negative association with Concern among men be found. Of further note, internal consistencies of NPI scale scores were mostly unacceptably low (Cronbach's α : .74 for Authority; .49 for Superiority, .66 for Exhibitionism, .45 for Exploitativeness, .71 for Vanity, and .48 for Entitlement).

Predicted null correlations were found for only five out of the 12 measures. Conscientiousness showed a weak negative association with Sensitivity among women, whereas it showed extraversion with Concern among men. Interestingly, openness overall showed small to medium-sized associations for both sexes which were stronger for Sensitivity than for Concern. Self-deceptive positivity was negatively associated with Sensitivity and Concern among men, whilst relationship exclusivity was a weak positive correlate of Concern. Among women, sexual attractiveness was a weak positive correlate of Sensitivity and Concern, as was erotophilic disposition of Sensitivity.

In order to explore the predictive value of the observed associations systematically, hierarchical regression analyses were performed on E-scale total, Sensitivity, and Concern scores, entering in a first step all variables that were found to correlate with the respective empathy scores (see Table 4, shaded fields), and they were done separately for men and women. For men, predictors remaining significant after backward elimination for Sensitivity were (in descending order of the magnitude of associations) openness, self-deceptive positivity, interpersonal manipulation, age, and neuroticism; whereas for Concern the predictors were neuroticism, interpersonal manipulation, openness, emotional investment, agreeableness, and social dominance orientation (Table 5). For total scores, extraversion emerged as significant predictor as well. Otherwise, predictors were a mix of those for the two scales. Predictors explained 20% (Sensitivity) up to 30% (total score) of variance.

For women, significant predictors of Sensitivity were emotional investment, openness, less prescriptive sex-role beliefs, conscientiousness, and nonsexual violence; whereas predictors of Concern were, social dominance orientation, emotional investment, nonsexual violence, and interpersonal manipulation (Table 6). For total scores, age emerged as a significant predictor, too. Predictors explained 18% (Sensitivity) up to 22% (total score) of variance.

TABLE 5
HIERARCHICAL REGRESSION ANALYSES: MEN

Measure	Total Score		Sensitivity		Concern	
	β	<i>p</i>	β	<i>p</i>	β	<i>p</i>
Age	-.13	.004	-.16	.001		
Agreeableness	.15	.006			.14	.012
Neuroticism	.20	< .001	.11	.033	.34	< .001
Extraversion	-.15	.004				
Openness	.30	< .001	.35	< .001	.15	.002
Emotional investment	.14	.004			.15	.002
Social dominance orientation	-.10	.040			-.12	.016
Interpersonal manipulation	-.13	.008	-.12	.014	-.16	.002
Self-deceptive positivity	-.12	.019	-.20	< .001		
<i>R</i>	.56		.46		.54	
Adjusted <i>R</i> ²	.297		.203		.277	
<i>F</i> ^a	18.73		20.35		21.56	

Note.—^a*df*₁ = 9, *df*₂ = 368 for total score, 5 and 374 for Sensitivity, and 7 and 369 for Concern, respectively, *ps* < .001.

DISCUSSION

This structural analysis of the E-scale for the most part supported previous findings, in that this measure assesses a general factor of empathy as well as of its four facets, namely factorial combinations of empathy reality statuses (real-life vs. fictitious situations) and empathy components (cognitive vs. emotional). However, in view of the present results, based on methods specifically suited to the ordered categorical item format of the E-scale, one prior assumption regarding the mutual orthogonality of these facets seems in need of revision. In particular, the four facets seem to reflect more strongly differences in empathy reality status than in empathy components.

Thus, the study at hand confirms that emotional and cognitive empathy are indeed indispensable parts of the same construct and may also be measured adequately with the E-scale. However, at the same time, these domains of empathy may not be structurally differentiated—at least in samples of healthy subjects as investigated here. This result in turn suggests that the distinction of cognitive and emotional empathy, even though it was of major interest in past research, appears to lack sufficient empirical justification. This appears not only true with regard to the E-scale, as investigated here, but also with regard to other multidimensional self-report measures (e.g., the Interpersonal Reactivity Index and the Empathy Quotient). However, the extent of overlap of cognitive and emotional

TABLE 6
HIERARCHICAL REGRESSION ANALYSES: WOMEN

Measure	Total Score		Sensitivity		Concern	
	β	p	β	p	β	p
Age	-.13	.016				
Openness	.15	.007	.19	.001		
Conscientiousness			-.17	.002		
Less prescriptive sex-role beliefs			.19	.001		
Emotional investment	.23	< .001	.22	< .001	.28	< .001
Social dominance orientation	-.23	< .001			-.29	< .001
Interpersonal manipulation					-.14	.008
Nonsexual violence	.16	.002	.11	.033	.15	.004
R	.48		.44		.48	
Adjusted R^2	.221		.179		.216	
F^a	18.14		14.20		21.75	

Note.—^a $df_1 = 5$, $df_2 = 297$ for total score and Sensitivity, and 4 and 298 for Concern, respectively, $ps < .001$.

empathy found here may be exceptionally high (factor correlation of .94 in Model 4). In contrast, differentiating systematically with regards to reality status—which is conceptually neglected in operationalizations of other available measures—appeared to better capture the structure of the E-scale than differentiating between emotional and cognitive components, and consequently also allowed some interesting insights with respect to the discriminant and concurrent validity of empathy, which are discussed in detail below.

Regarding the primary hypothesis, respondent sex proved an important correlate of empathy (women reported noticeably higher levels of empathy than men), explaining 6% (Sensitivity) to 13% (Concern) of the variance. Notwithstanding this evidence, sex differences in real-life empathy, i.e., not based on self-reports, may be difficult to ascertain (e.g., Ickes, Gesn, & Graham, 2000). Yet, correlational analysis of a variety of assumed convergent constructs of the E-scale revealed further sex-dependent patterns of associations. Less prescriptive views on the roles of men and women and emotional investment in a relationship were positive correlates of total and scale scores across both sexes. Also, some relationship-related scales, presumed to be independent of empathy, were found to be differentially associated with E-scale scores for men vs. women. Although most of these associations were weak, empathy, as measured by the E-scale, seemed to be consistently predicted by a variety of gender-related and relationship-related traits.

Social dominance orientation, Machiavellianism, and interpersonal manipulation were the strongest negative correlates of empathy across both sexes. Negative associations of Machiavellianism with empathy have already been reported previously (e.g., Andrew, Cooke, & Muncer, 2008). However, the distinction of empathy arising from fictitious vs. real-life situations proved most interesting here as well, as it revealed that these variables impacted real-life empathy more strongly or even exclusively (as was the case for Machiavellianism).

Unexpectedly, impulsive thrill-seeking (a component of psychopathy, as measured with the SRP-III) and nonsexual violence were positive (but weak) correlates of Sensitivity among women. Given that (1) women scored considerably lower on psychopathy and its components than men and also scored only slightly higher on nonsexual violence than men (which is consistent with previous results; see Dobash, *et al.*, 2000), (2) associations with nonsexual violence were only limited to minor (and only verbal) acts of violence towards partners, and (3) in light of the observed associations with less prescriptive sex-role beliefs as well as with emotional investment, we suggest that these slight positive correlations do not necessarily reflect "pathological" tendencies. Instead, these may well be linked to tendencies of personal (healthy) self-liberation which still seem especially important to women and which may also be addressed by movies and texts. This assumption should be followed up in further research, yet it has to be emphasized that observed associations were only weak.

Narcissism, as measured with the NPI (the 40 items of which follow the eight criteria for narcissistic personality disorder of the DSM-III), failed to show substantial associations with the E-scale. In contrast, Watson, *et al.* (1984) reported associations of the NPI with the HES and the QMEE. We conclude that this may be at least partly due to construct-irrelevant contents in those measures. First, only one of the eight criteria of narcissistic personality disorder concerns the lack of empathy. Hence, total NPI scores in all likelihood do not solely reflect lack of empathy. Second, the ISDP-2 respondents, whose data were analyzed here, constituted a general population sample (not psychiatrically screened or clinically diagnosed with narcissistic personality disorder). Expected effects of narcissism on E-scale scores may therefore be more easily obtained in comparisons of clinical vs. non-clinical samples (note, however, that the NPI was designed to measure narcissism in the general population). However, third and most importantly, the NPI has been repeatedly reported to have an unclear factor structure and to be only a weak measure of its intended construct (for a recent overview, see Ackerman, Witt, Donnellan, Trzesniewski, Robins, & Kashy, 2011; we also tested the additional scales suggested by these researchers, but arrived exactly at comparable

null associations). The reliabilities of NPI scales were also mostly unacceptably low in the present study. Hence, we suggest that, rather than being a shortcoming of the E-scale regarding its concurrent validity, this was primarily caused by using the NPI. Further research is needed on this subject, though. As opposed to prior evidence (Leibetseder, *et al.*, 2007), self-esteem was also not strongly related to empathy. This difference may have arisen through different self-esteem scales used.

For both sexes, the Big Five personality dimensions were found to be differentially predictive of empathy. Most notably, openness was considerably positively correlated with E-scale scores, albeit stronger with Sensitivity than with Concern. Empathy may be construed as the sensitivity to notice and understand emotional states of others, but also as the willingness or ability to think about and to show some openness to such states. Based on the current evidence, openness appears to have an important role for the formation of empathic responses, which has not been highlighted by prior related research. Neuroticism was a correlate of empathy only among men, as was the case for extraversion. Among men, agreeableness was a stronger correlate of empathy, whereas among women agreeableness correlated only with Concern.

E-scale score variance could be predicted to a larger extent for men (20% to 30%) than for women (18% to 22%) in our sample. Given women's higher scores, the difference in explained variance may be due to a ceiling effect, restricting observable associations due to distributional truncation. Yet, regression analyses also highlighted differential associations of other constructs with Sensitivity vs. Concern in men vs. women: whereas among men openness was the single most dominant predictor of Sensitivity, among women openness, conscientiousness, less prescriptive sex-role beliefs, and emotional investment contributed equally to the prediction of Sensitivity. Note that, unexpectedly, self-deceptive positivity was the second-largest predictor here in men; overconfidence in one's judgements and rationality thus seemed reciprocally associated with men's empathic reactions to movies and texts. Women's slight association of impression management (which was assumed a convergent construct for both sexes) did not survive the elimination procedure in the regression analysis. In contrast, neuroticism dominated prediction of Concern among men, whilst emotional investment and social dominance orientation were most important and equally predictive among women.

These results highlight that different traits and attitudes influence self-reports of empathy among men and women in this sample; furthermore, that the distinction along reality status proved meaningful here as well. We thus believe that in further research this distinction may also help to sort out and to clarify previously faced problems with the construct validity of self-reported empathy.

Limitations and Conclusions

Limitations of this study pertain to assessing empathy in the course of a larger survey on personality and sexuality, which may have triggered or amplified sex differences with regard to empathy. Moreover, sampling covered only urban areas which possibly also affected results. Furthermore, investigation of a large number of variables in the correlational analyses may have resulted in Type I errors. However, with regard to this and the mostly small observed associations, regression analyses confirmed the practical significance of many of the investigated variables, as percentages of explained E-scale score variances were sizable.

In conclusion, empathy, as measured by the E-scale, is related to demographic sex, gender-related and relationship-related variables, anti-social attitudes and behavior, and several of the Big Five dimensions. The structure of the E-scale was found to be stable and replicable. Regarding its distinction between the components of emotional vs. cognitive empathy on the one hand, and empathy arising from real-life situations vs. from fictitious situations on the other hand, we found that the former could not actually be discerned empirically, as has also been reported for related multidimensional measures of empathy, whereas the latter distinction proved to be an interesting one, the significance of which was highlighted in this study and should be followed up by future research. Further investigations should also address some unexpected associations of Sensitivity with several constructs among women (nonsexual violence, impulsive thrill-seeking) and men (self-deceptive positivity), as observed in the current data. Notwithstanding these remaining agenda for future inquiry, applications of the E-scale—as a psychometrically sound and in-depth, but still economic, self-report measure of empathy—for research into personality and other individual differences variables seem to have promise from these results.

REFERENCES

- ACKERMAN, R. A., WITT, E. A., DONNELLAN, M. B., TRZESNIEWSKI, K. H., ROBINS, R. W., & KASHY, D. A. (2011) What does the Narcissistic Personality Inventory really measure? *Assessment*, 18, 67-87.
- AMERICAN PSYCHIATRIC ASSOCIATION (APA). (2000) *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV-TR*. (4th ed., text revision) Washington, DC: American Psychiatric Association.
- ANDREW, J., COOKE, M., & MUNCER, S. J. (2008) The relationship between empathy and Machiavellianism: an alternative to empathizing-systemizing theory. *Personality and Individual Differences*, 44, 1203-1211.
- BARON-COHEN, S., & WHEELWRIGHT, S. (2004) The Empathy Quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163-175.
- BRADBURN, N. M. (1969) *The structure of psychological well-being*. Chicago, IL: Aldine.

- BRUNNER, M., NAGY, G., & WILHELMS, O. (2012) A tutorial on hierarchically structured constructs. *Journal of Personality*, 80, 796-846.
- BRYANT, B. K. (1982) An index of empathy for children and adolescents. *Child Development*, 53, 413-425.
- CHRISTIE, R., & GEIS, F. L. (1970) *Studies in Machiavellianism*. New York: Academic Press.
- CLIFFORDSON, C. (2001) Parents' judgements and students' self-judgements of empathy: the structure of empathy and agreement of judgements based on the Interpersonal Reactivity Index (IRI). *European Journal of Psychological Assessment*, 17, 36-47.
- CLIFFORDSON, C. (2002) The hierarchical structure of empathy: dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology*, 43, 49-59.
- COHEN, D., & STRAYER, J. (1996) Empathy in conduct-disordered and comparison youth. *Developmental Psychology*, 32, 988-998.
- DAVIS, M. H. (1983) Measuring individual differences in empathy: evidence from a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- DOBASH, R. E., DOBASH, R. P., CAVANAGH, K., & LEWIS, R. (2000) *Changing violent men*. Thousand Oaks, CA: Sage.
- EISENBERG, N., & LENNON, R. (1983) Sex differences in empathy and related capacities. *Psychological Bulletin*, 94, 100-131.
- EISENBERG, N., & MILLER, P. (1987) Empathy and prosocial behavior. *Psychological Bulletin*, 101, 91-119.
- EYSENCK, S. B. G., & EYSENCK, H. J. (1978) Impulsiveness and venturesomeness: their position in a dimensional system of personality description. *Psychological Reports*, 43, 1247-1253.
- FLETCHER, G. J. O., SIMPSON, J. A., & THOMAS, G. (2000) The measurement of perceived relationship quality components: a confirmatory factor analytic approach. *Personality and Social Psychology Bulletin*, 26, 340-354.
- GEER, J. H., ESTUPINAN, L. A., & MANGUNO-MIRE, G. M. (2000) Empathy, social skills, and other relevant cognitive processes in rapists and child molesters. *Aggression and Violent Behavior*, 5, 99-126.
- GRAZIANO, W. G., & EISENBERG, N. (1997) Agreeableness: a dimension of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology*. San Diego, CA: Academic Press. Pp. 795-824.
- HOGAN, R. (1969) Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307-316.
- HOLZ-EBELING, F., & STEINMETZ, M. (1995) [How useful are the available empathy questionnaires? Critical analyses taking item content into consideration]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 16, 11-32.
- HU, L., & BENTLER, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- ICKES, W., GESN, P. R., & GRAHAM, T. (2000) Gender differences in empathic accuracy: differential stability or differential motivation? *Personal Relationships*, 7, 95-109.
- JOHN, O. P., & SRIVASTAVA, S. (1990) The Big Five trait taxonomy: history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: theory and research*. New York: Guilford. Pp. 102-138.
- JOLLIFFE, D., & FARRINGTON, D. P. (2004) Empathy and offending: a systematic review and meta-analysis. *Aggression and Violent Behavior*, 9, 441-476.

- KALIN, R., & TILBY, P. J. (1978) Development and validation of a sex-role ideology scale. *Psychological Reports*, 42, 731-738.
- KLINE, M. (1992) *The handbook of psychological testing*. London: Routledge.
- LAWRENCE, E. J., SHAW, P., BAKER, D., BARON-COHEN, S., & DAVID, A. S. (2004) Measuring empathy: reliability and validity of the empathy quotient. *Psychological Medicine*, 34, 911-924.
- LEIBTSEDER, M., LAIREITER, A.-R., & KÖLLER, T. (2007) Structural analysis of the E-scale. *Personality and Individual Differences*, 42, 547-561.
- LEIBTSEDER, M., LAIREITER, A.-R., RIEPLER, A., & KÖLLER, T. (2001) [E-scale: questionnaire for the assessment of empathy—description and psychometric properties]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22, 70-85.
- LOTTE, I. L. (1988) Rape Supportive Attitude Scale. In C. M. Davis, W. L. Yarber, & S. L. Davis (Eds.), *Handbook of sexuality-related measures*. Thousand Oaks, CA: Sage. Pp. 235-237.
- MEHRABIAN, A., & EPSTEIN, N. (1972) A measure of emotional empathy. *Journal of Personality*, 40, 525-543.
- MILLER, P. A., & EISENBERG, N. (1988) The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological Bulletin*, 103, 324-344.
- MOSHER, D. L. (1998) Aggressive Sexual Behavior Inventory. In C. M. Davis, W. L. Yarber, R. Bauserman, G. Schreer, & S. L. Davis (Eds.), *Handbook of sexuality-related measures*. Thousand Oaks, CA: Sage. Pp. 55-56.
- MUNCER, S. J., & LING, J. (2006) Psychometric analysis of the empathy quotient (EQ) scale. *Personality and Individual Differences*, 40, 1111-1119.
- MUTHÉN, L. K., & MUTHÉN, B. O. (2008) *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- NETTLE, D. (2007) Empathizing and systemizing: what are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology*, 98, 237-255.
- PAULHUS, D. L. (1991) Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press. Pp. 17-59.
- PRATTO, F., SIDANIUS, J., STALLWORTH, L. M., & MALLE, B. F. (1994) Social dominance orientation: a personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67, 741-763.
- PREMACK, D., & WOODRUFF, G. (1978) Does the chimpanzee have a 'theory-of-mind'? *Behavioral and Brain Sciences*, 4, 515-526.
- PULOS, S., ELISON, J., & LENNON, C. (2004) The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality*, 32, 355-360.
- RASKIN, R., & HALL, C. S. (1979) A narcissistic personality inventory. *Psychological Reports*, 45, 590.
- RENIERS, R. L. E. P., CORCORAN, R., DRAKE, R., SHRYANE, N. M., & VÖLLM, B. A. (2011) The QCAE: a questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, 93, 84-95.
- ROGERS, K., DZIOBEK, I., HASSENSTAB, J., WOLF, O. T., & CONVIT, A. (2007) Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders*, 37, 709-715.

- ROSENBERG, M. (1965) *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- SCHMITT, D. P., ALCALAY, L., ALLENSWORTH, M., ALLIK, J., AULT, L., AUSTERS, I., . . . ZUPANËIË, A. (2003) Are men universally more dismissing than women? Gender differences in romantic attachment across 62 cultural regions. *Personal Relationships*, 10, 307-331.
- SCHMITT, D. P., ALCALAY, L., ALLENSWORTH, M., ALLIK, J., AULT, L., AUSTERS, I., . . . ZUPANËIË, A. (2004) Patterns and universals of adult romantic attachment across 62 cultural regions: are models of self and of other pancultural constructs? *Journal of Cross-Cultural Psychology*, 35, 367-402.
- SCHMITT, D. P., ALLIK, J., MCCRAE, R. R., & BENET-MARTINEZ, V. (2007) The geographic distribution of Big Five personality traits: patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38, 173-212.
- SCHMITT, D. P., & BUSS, D. M. (2000) Sexual dimension of person description: beyond or subsumed by the Big Five? *Journal of Research in Personality*, 34, 141-177.
- SCHMITT, D. P., REALO, A., VORACEK, M., & ALLIK, J. (2008) Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94, 168-182.
- SHAPIRO, J., & RUCKER, L. (2004) The Don Quixote effect: why going to the movies can help develop empathy and altruism in medical students and residents. *Families, Systems, and Health*, 22, 445-452.
- SIMPSON, J. A., & GANGESTAD, S. W. (1991) Individual differences in sociosexuality: evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology*, 60, 870-883.
- SPRENG, R. N., MCKINNON, M. C., MAR, R. A., & LEVINE, B. (2009) The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution of multiple empathy measures. *Journal of Personality Assessment*, 91, 62-71.
- STEINS, G. (1998) [Diagnosis of empathy and perspective-taking: a review of the relation between both constructs and implications for their assessment]. *Diagnostica*, 44, 117-129.
- STOTLAND, E., KENNETH, E., MATHEWS, J., SHERMAN, S. E., HANSSON, R. O., & RICHARDSON, B. (1978) Empathy, fantasy and helping. *Sage Library of Social Research*, 65, 27-45.
- WATSON, P. J., GRISHAM, S. O., TROTTER, M. V., & BIDERMAN, M. D. (1984) Narcissism and empathy: validity evidence for the Narcissistic Personality Inventory. *Journal of Personality Assessment*, 48, 301-305.
- WILLIAMS, K. M., NATHANSON, C., & PAULHUS, D. L. (2003) Structure and validity of the Self-Report Psychopathy Scale-III in normal populations. Poster presented at the 111th Annual Convention of the American Psychological Association, Toronto, Canada, August 7-10, 2003.

Accepted August 31, 2013.

GENERALIZABILITY OF DELAY OF GRATIFICATION: DIMENSIONALITY AND FUNCTION¹

XIAOYAN LIU, LEI WANG, AND JING JIANG

*Department of Psychology
Beijing University*

Summary.—There is a debate about the factor structure of adults' ability to delay gratification and mixed findings concerning the relationship between delay of gratification and achievement. Three studies were conducted to show that delay of gratification had two components. In Study 1, exploratory factor analyses showed that the Generalizability of Deferment of Gratification Questionnaire had two factors: Controlling-Impulse and Planning-and-Waiting. Study 2 verified the two-factor structure by confirmatory factor analysis and demonstrated acceptable reliability, construct and divergent validity. Specifically, Planning-and-Waiting was correlated with delay-discounting, self-control, uncertainty avoidance, Openness, Conscientiousness, Agreeableness, and self-efficacy, whereas Controlling-Impulse was correlated with self-control, Conscientiousness, and Agreeableness. Moreover, Planning-and-Waiting was the unique predictor of CFC-Future, but Controlling-Impulse predicted substantive variance in both CFC-Future and CFC-Immediate. Study 3, using multi-wave and multi-source data, further showed that only Controlling-Impulse was an important predictor of long-term performance and creative performance, supporting the distinctiveness of the two factors.

Delay of gratification is a choice orientation in which individuals try to forego an immediate gratification to attain a more valuable outcome later on (Mischel, 1974). Many studies have documented the importance and implications of the ability to delay gratification in early childhood for lifelong development (see Ayduk, 2007, for review). As a kind of self-regulation, previous research even indicates that delay of gratification has a bigger effect on academic performance than IQ does, because delay of gratification may help students display more academic engagement behavior (Duckworth & Seligman, 2005).

Although delay of gratification is so important, its structure is still not clear. The experimental paradigm of delay of gratification adopted in most research merely measures the length of time a child can wait for a delayed, larger reward (Mischel, 1974); but does not measure the components of delay of gratification. Moreover, all delay of gratification questionnaires (e.g., GDGQ, Ray & Najman, 1986; ADGS, Bembenuddy, & Karabenick, 1998) consider delay of gratification to be unidimensional. Although Mischel (1974) proposed a two-phase model of delay of gratification—

¹Address correspondence to Prof. Lei Wang, Ph.D., Department of Psychology, Peking University, Beijing, China or e-mail (leiwang@pku.edu.cn).

abandoning enjoyment and maintaining goals—this has never been supported in empirical research. There are inconsistent results regarding the relation between delay of gratification and other constructs in previous research. For instance, delay-discounting is often used as an equivalent of delay of gratification (Reynolds, De Wit, & Richards, 2002), but the relation between the two variables is unclear (Kirby, Winston, & Santiesteban, 2005). This might be due to multiple components in delay-of-gratification questionnaires and their different functions. A multi-dimensional questionnaire might help explain why there has been inconsistency in reports of delay of gratification predicting other behaviors.

The current research will address the above questions by employing exploratory and confirmatory factor analysis, and re-examining the validity and function of delay of gratification. Specifically, Ray and Najman's (1986) Generalizability of Deferment of Gratification Questionnaire (GDGQ) will be used. Apparently, the GDGQ is the only measure of the general personality trait rather than delay of gratification in a specific situation (e.g., Ward, Perry, Woltz, & Doolin, 1989; Bembenuitty & Karabenick, 1998). In addition, GDGQ has been widely accepted and used; there is a good deal of research investigating how delay of gratification relates to various behaviors and traits, such as academic behavior (Bembenuitty & Karabenick, 1998), life satisfaction (Caldwell & Mowrer, 1998), organizational commitment and job satisfaction (Witt, 1990a), consideration of future consequences (Strathman, Gleicher, Boninger, & Edwards, 1994), gambling behavior (Parke, Griffiths, & Irwing, 2004), social responsibility (Witt, 1990b), and debt (Norvilitis & MacLean, 2006). For these reasons, the GDGQ was adopted to explore the structure of delay of gratification.

Factor Structure of Delay of Gratification

Delay of gratification may have a two-factor structure, based on both theoretical claims and semantic analysis of the GDGQ's items. From a theoretical viewpoint, Mischel (1974) proposed a two-phase model for delay of gratification. During phase 1, "delayers" would make a choice to abandon immediate gratification for the sake of a delayed but more valuable outcome. During phase 2, "delayers" would maintain their choice until the eventual goal was achieved. (For example, in Mischel's experiment, children had to overcome the immediate temptation for a piece of cookie and wait 15 minutes for two cookies.) Delay of gratification may have two factors: one is abandoning the immediate gratification, and the other is maintaining the choice or planning for the larger future reward. The semantics of the GDGQ (Ray & Najman, 1986) indicate that items may represent two different aspects of delayed gratification. For instance, the items like "Would you describe yourself as often being too impulsive for your own good" may assess whether individuals can control impulses.

Other items, such as “I enjoy a thing all the more because I have to wait for it or plan for it,” aim to assess whether individuals could plan in advance and wait for a long time.

Metcalf and Mischel (1999) proposed a two-system framework in the processing of delay of gratification, a cognitive “cool” system and an emotional “hot” system, a mechanism underlying anger induction, expression, and regulation (Lok, Bond, & Tse, 2009). When the hot impulse for immediate but smaller value reward is triggered, the cool system will control impulse as well as strategically plan for the larger reward. According to this processing perspective, delay of gratification may have a two-factor structure, quite similar to that predicted from Mischel's theory.

Convergent Validity

Convergent validity represents the extent to which a scale relates to other measures of the same or similar constructs (Hinkin, 1998). We adopted delay-discounting, self-control, and consideration of future consequence as convergent criteria which are always regarded as having overlap with delay of gratification in theoretical constructs (Reynolds, *et al.*, 2002; Petrocelli, 2003).

Delay-discounting refers to “the reduction in the present value of a future reward as the delay to that reward increases” (Kirby, Petry, & Bickel, 1999; pp. 78), often viewed as reflecting processes equivalent to delay of gratification (Rachlin, Brown, & Cross, 2000). The more remote a future reward is, the lower the present value it has, and, therefore, the less likely the reward is to be chosen among current alternatives (Kirby, *et al.*, 1999). Researchers have pointed out that delay-discounting is related to time perception and future orientation (Schweizer, 2002; Scholten & Read, 2006; Zauberman, Kim, Malkoc, & Bettman, 2009) than to impulse control (Steinberg, Graham, O'Brien, Woolard, Cauffman, & Banich, 2009). Thus, it could be expected that a planning-and-waiting factor, rather than a controlling-impulse factor, would be negatively correlated with the delay-discounting rate.

Self-control is the ability to override or change one's inner responses, as well as to interrupt undesired behavioral tendencies and refrain from acting on them (Tangney, Baumeister, & Boone, 2004). In recent decades, the experimental paradigms of delay of gratification have always been used as a way to measure one's self-control ability in laboratory settings (Mischel, 1961, 1974; Mischel, Shoda, & Rodriguez, 1989), and the inability to delay gratification is often considered to be caused by a loss of self-control (Muraven & Baumeister, 2000). Previous research also indicated that most acts of self-control involved both sacrificing short-term happiness and achieving long-term well-being (Rachlin, 2000). Therefore, self-control is likely related to both planning-and-waiting and controlling-impulse factors.

When people decide to delay gratification, they must consider immediate vs. future consequences. Past research has documented a positive relation between delay of gratification and consideration of future consequences; the latter refers to the extent to which individuals consider the potential distant outcomes of their current behaviors and the extent to which they are influenced by these potential outcomes (Strathman, *et al.*, 1994; p. 743). The Consideration of Future Consequences scale (CFC scale) comprises two underlying subfactors: concern with immediate consequences (CFC-Immediate) and future consequences (CFC-Future), respectively (Petrocelli, 2003; Joireman, Balliet, Sprott, Spangenberg, & Schultz, 2008). Because planning-and-waiting focuses on delayers' insistence on waiting for a more valuable reward in the future, it may be more closely related to CFC-Future. Controlling impulses focuses on delayers making the choice to abandon immediate gratification, so a controlling-impulse factor may be more closely related to CFC-Immediate.

Criterion-related Validity

Criterion-related validity, or the extent to which a construct is related to variables derived from theory, is an important aspect of construct validity (Hinkin, 1998). We chose the measures Consideration of Future Consequences, three dimensions of a Big Five personality measure (Conscientiousness, Agreeableness, and Openness), uncertainty avoidance, and self-efficacy as validity criterion measures.

Big Five personality measures are among the most widely used personality assessments. Some scores have been found to be related to delay of gratification behavior. Krueger, Caspi, Moffitt, White, & Stouthammer-Loeber (1996) found that 12- and 13-year-old boys who more often chose the delayed option were described as conscientious, agreeable, and open to experience by their mothers. Because the ability to delay gratification is a competency that may be predictable from multiple aspects of personality, Openness, Conscientiousness, and Agreeableness were adopted as criterion variables. The two delay-of-gratification factors should have different correlations with these three personalities.

Uncertainty avoidance represents the extent to which the individuals feel threatened by uncertain or unknown situations (Hofstede, 1991). Previous research has indicated that one of the effective ways to avoid uncertainty is to plan carefully (Rauch, Frese, & Sonnentag, 2000). We also infer that when a person plans for a future larger reward or promotion at work, he also is planning to avoid uncertainty. Therefore, uncertainty avoidance should be related to the planning-and-waiting factor of delay of gratification; in contrast, there may be little relationship between uncertainty avoidance and the controlling-impulse factor.

In previous research, it has been found that students reporting greater delay of gratification were higher in self-efficacy (Bembenutty & Karabenick, 1998), and adults' self-efficacy was associated positively with their actual delay behaviors, such as dieting (Rosenbaum & Ben-Ari Smira, 1986). Self-efficacy refers to the belief about being able to control challenging environmental demands by means of taking adaptive action (Bandura, 1977). People higher in self-efficacy are more confident that they will have a positive future (Kerpelman & Mosher, 2004). For these reasons, it can be predicted that self-efficacy would be more closely related with the planning-and-waiting than the controlling-impulse factors of delay of gratification.

Functions of Different Delay of Gratification Factors

Previous research has shown that delay of gratification could predict people's performance over a long time. For example, 4-year-old children who waited longer in delay of gratification task achieved higher SAT scores as adolescents (Mischel, Shoda, & Peake, 1988; Mischel, *et al.*, 1989). However, only cross-sectional studies have shown that delay of gratification was correlated positively with employees' job performance (Miller, Woehr, & Hudspeth, 2002). No study has demonstrated empirically that delay of gratification can predict job performance over a period of time. In the current study, longitudinal data were gathered to assess the causal relationship between delay of gratification and job performance. Specifically, a multi-dimensional delay of gratification questionnaire was used to identify which component of delay of gratification might predict job performance. At the same time, this study could be further evidence of the two-component construct of delay of gratification.

The expected relation between delay of gratification and job performance is based on two lines of logic. Firstly, Mischel's early research (Mischel, 1981; Mischel & Mischel, 1983) found that effective delay in the children's waiting paradigm included the ability to control and divert their attention from temptation flexibly and distract themselves from the excessive arousal of short-term outcomes purposefully; both are strategies of controlling impulses. Secondly, in academic performance Bembenutty and Karabenick (1998) showed that students with higher scores on academic delay of gratification could postpone "the immediately available opportunities to satisfy impulses in favor of pursuing important but remote academic rewards or goals." Based on these previous findings, it can be inferred that the controlling-impulse factor would be more closely related to adults' job performance than a planning-and-waiting factor.

Apart from job performance, the roles of the two delay-of-gratification factors on creative performance were examined. Creativity in the

employment arena is the production of novel and useful ideas by employees—idea that can be the starting points for innovation (Oldham & Cummings, 1996). Creativity in the current samples of high-tech company employees should be an indicator of performance, because when employees exhibit creativity at work they produce novel, potentially useful ideas about organizational products, practices, services, or procedures (Shalley & Zhou, 2008). Therefore, it was predicted that the relations between the two delay-of-gratification factors and creative performance would be similar to their relations with job performance.

In the current study, the possible two-factor structure of the GDGQ was first examined using exploratory factor analysis in Study 1. In Study 2, the GDGQ's two-factor structure was verified by confirmatory factor analyses and its validity was examined. In Study 3, the roles of the two delay-of-gratification factors in predicting job and creative performance were explored.

Study 1

Study 1 explored the possible factor structure of GDGQ. For research purposes, adult samples were recruited from organizations and an Exploratory Factor Analysis (EFA) of the GDGQ's 12 items was conducted to obtain a preliminary view of the overall relational structure of these items.

METHOD

Participants

Participants were recruited from four information technology companies in China. A total of 322 applicants (190 men) were asked to complete the GDGQ voluntarily. The average age was 27.2 yr. ($SD = 4.0$). Among these participants, 98.4% had a college diploma or higher degree.

Measures

The 12-item Generalizability of Deferment of Gratification Questionnaire (GDGQ; Ray & Najman, 1986) was used to measure general delay of gratification. The respondents rated to what extent they agreed with each item on a 7-point scale with anchors 1: Strongly disagree and 7: Strongly agree.

Translations

The GDGQ was translated and back-translated using approved techniques (Bracken & Barona, 1991). Firstly, a researcher whose native language was Chinese translated the scales from English to Chinese. Then, a Chinese translator majoring in English who did not know the study's purpose completed the back-translation. Thirdly, another researcher compared the two English versions and checked whether each item's mean-

ing was equivalent or not. Finally, any differences were discussed and a final translated version was agreed upon by the two researchers and the translator.

RESULTS

Before Exploratory Factor Analysis, the scores of 6 items were reversed in the GDGQ (e.g., "I like to spend my money as soon as I get it"). The KMO measure of sampling adequacy (.74) and the Bartlett test of sphericity (569.62, $p < .001$) demonstrated adequate multivariate normality. EFA was conducted using SPSS Version 18.0. Principal component analysis (PCA) was performed on the 12 items. A two-factor solution was finalized based on the criteria: factor eigenvalues greater than 1, interpretability of factors, and at least 3 strongly loading items per factor (DeCoster, 1998). In addition, the scree plot suggested that the absolute slope showed little decrease after two factors. Item 10 was eliminated because its loadings were less than .30 on both factors. Although the cumulative variance of the two factors was not very high (39.58%), extracting three factors would have violated the principle that the number of items would be less than three in subsequent factors (Kaiser, 1974). The first factor was labeled Controlling-Impulse, since these items represent not succumbing to immediate gratification. The second factor was labeled Planning-and-Waiting, because these items suggest a willingness to wait for a possibly larger reward and plan for the future (see Table 1).

Study 2

Confirmatory factor analysis (CFA) was run with AMOS Version 7.0 to confirm the two-factor structure obtained in Study 1, with all GDGQ items loading on two latent factors. In this study, the convergent and criterion-related validity of the two-factor GDGQ also were examined.

METHOD

Participants and Procedure

Sample 1.—A new sample of Chinese participants (129 men, 193 women) were recruited from various occupations using a snowball sampling technique via e-mails containing study information, the link to the survey, and instructions to forward the e-mail to friends and colleagues. The respondents voluntarily filled out the questionnaires online. They had a mean age of 28.1 yr. ($SD = 6.80$), and 76.0% had a college diploma or higher degree. The GDGQ, uncertainty avoidance, and the Consideration of Future Consequences scale were administered (see below).

Sample 2.—Another sample of 165 Chinese people (69 men, 96 women; M age = 27.1 yr., $SD = 3.6$) were recruited from an auto dealership in a sales training course. Among these participants, 90.9% had a college diploma or

TABLE 1
FINAL VERSION OF GENERALIZABILITY OF DEFERMENT OF GRATIFICATION QUESTIONNAIRE
IN STUDY 1 (N = 322)

Scale Item	Component Loading		Communality
	CI	PW	
9. 我喜欢一有钱就扣它们花掉 [I like to spend my money as soon as I get it.] (R)	.78	.02	.60
5. 我经常“财政透支” [I am constantly “broke.”] (R)	.64	.18	.44
6. 我愿意这样的观点“吃喝玩乐吧，因为明天我们可能都死了” [I agree with the philosophy: “Eat, drink and be merry, for tomorrow we may be all dead.”] (R)	.60	.14	.38
4. 当我在超市的时候，会买很多原来没有计划买的东西 [When I am in a supermarket, I always tend to buy a lot of things I hadn't planned to buy.] (R)	.57	.10	.34
7. 我经常一时冲动地去做一些慷慨之事 [I describe myself as often being too impulsive for my own good.] (R)	.55	.10	.36
3. 当我是个小孩儿的时候，我就试图攒下我的零花钱 [I tended to save my pocket money as a child.]	.54	-.26	.32
1. 我一般能做到不立刻扣钱花掉，而是扣钱省下夹 [I am good at saving my money rather than spending it straight away.]	.51	.25	.32
2. 如果为了得到一件东西我必须做出计划和等待，那么我会更加享受这件东西 [I enjoy a thing all the more because I have had to wait for it and plan for it.]	.05	.69	.48
12. 我在超前计划方面做得很好 [I am good at planning things way in advance.]	.17	.66	.46
8. 我经常觉得在下决定之前的等待和思考都是值得的 [I fairly often find that it is worthwhile to wait and think things over before deciding.]	.15	.61	.40
11. 大多数情况下，即使处于被动的等待状态，我也能泰然处之 [I can tolerate being kept waiting for things fairly easily most of the time.]	.01	.50	.25
Variance explained (Total = 39.58%)	23.75	15.83	

Note.—CI = Controlling-Impulse; PW = Planning-and-Waiting. Boldface loadings are major factor loadings. (R) Item is reverse scored.

higher degree. The GDGQ, delay-discounting rate scale, three Big Five subscales (Agreeableness, Conscientiousness, Openness), and a self-control scale were administered (see below).

Measures

All of the following questionnaires' items were rated on a 7-point scale with anchors 1: Strongly disagree and 7: Strongly agree, except for the

delay discounting rate scale. Since this study used translated scales, CFA was run on each scale to ensure that the factor structure was the same as that of the original scales. Internal consistency reliabilities were checked as Cronbach's α and McDonald's ω (see Table 2 for complete information).

Uncertainty Avoidance Scale (Dorfman & Howell, 1988).—The 7-item Uncertainty Avoidance Scale was used to measure the tendency to avoid uncertainty. Example items are as follows: "Standard operating procedures are helpful to employees on the job." A higher score indicates the respondent's stronger tendency to avoid uncertainty. Internal consistency reliability of the scale was acceptable. The one-factor structure fit to the present data well: GFI = .98, AGFI = .96, CFI = .99, RMSEA = .051.

Consideration of Future Consequence Scale (CFC: Strathman, et al., 1994).—The 12-item scale was used to estimate the extent to which people consider the potential distant outcomes of their current behaviors and the extent to which they are influenced by these potential outcomes (Strathman, et al., 1994). The CFC has two subscales labeled CFC-Future (e.g., "I consider how things might be in the future, and try to influence those things with my day to day behavior") and CFC-Immediate (e.g., "My behavior is only influenced by the immediate") (Petrocelli, 2003; Joireman, et al., 2008). Higher scores on both the CFC-Total and CFC-Future scales reflect a higher concern with future consequences, whereas higher scores on the CFC-Immediate scale reflect a higher concern with immediate consequences. Internal consistency reliabilities of the scales were acceptable. The two-factor structure fit the present data acceptably: GFI = .94, AGFI = .88, CFI = .88, RMSEA = .082.

General Self-Efficacy Scale (GSE: Jerusalem & Schwarzer, 1992).—The 10-item scale measures the global confidence in one's coping ability across a wide range of stressful situations (e.g., "I am confident that I could deal efficiently with unexpected events"). High scores on the scale represent a more confident attitude towards stress-coping. Internal consistency reliability of the scale was good. The one-factor structure fit the present data acceptably: GFI = .97, AGFI = .92, CFI = .97, RMSEA = .076.

Delay-Discounting Rate Scale (Kirby & Marakovi, 1996).—The discount rate indicates the steepness of the reduction in present value with increases in delay (Kirby, et al., 1999). The higher the rate at which a person discounts future rewards, the lower the present values of future rewards are and the less effect those rewards will have on current choices (Kirby, et al., 1999). Kirby and Marakovi's (1996) monetary-choice questionnaire was used to estimate each participant's discount rate. The questionnaire included 27 questions and each question offered participants two options: an immediate reward and a delayed reward. Based on participants' choices of the immediate reward across 27 trials, a k value was calculated

TABLE 2
DESCRIPTIVE STATISTICS AND PEARSON CORRELATIONS AMONG VARIABLES FOR SAMPLE 1 IN STUDY 2 (N = 322)

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
			Correlations									
1. Gender	1.60	0.49										
2. Age	28.14	6.78	.01									
3. PW	4.63	0.93	-.13*	-.06	(.60/.62)							
4. CI	4.75	1.05	-.06	.16†	.18†	(.75/.70)						
5. Deferment of gratification	4.71	0.80	-.10	.11*	.57†	.91†	(.73/.72)					
6. CFC-Total	4.32	0.73	-.17†	-.12*	.23†	.26†	.31†	(.70/.78)				
7. CFC-Future	4.40	0.72	-.18†	-.05	.29†	.25†	.33†	.73†	(.72/.76)			
8. CFC-Immediate	4.26	0.94	.13*	.13*	-.14*	-.21†	-.26†	-.93†	-.42†	(.70/.74)		
9. Uncertainty avoidance	5.01	0.97	-.04	.13*	.36†	.10	.23†	.09	.27†	.03	(.70/.77)	
10. Self-efficacy	4.59	0.96	-.15*	-.05	.45†	.08	.25†	.17*	-.21†	-.11	.33†	(.86/.85)

Note.—Gender: men = 1; women = 2; CI = Controlling-Impulse; PW = Planning-and-Waiting. Cronbach's alphas and McDonald's omegas (α/ω) are on the diagonal in brackets. * $p < .05$. † $p < .01$.

according to a given formulation. k is a parameter that reflects the extent to which future rewards are diminished in value as a function of the delay that must be endured to receive them. The distributions of k s were approximately normalized using the natural log transformation, so the following calculations were based on $\text{Ln}k$ (see Kirby, *et al.*, 1999, for review). Higher k and $\text{Ln}k$ indicate participants' stronger tendency to choose an immediate reward.

Agreeableness, Conscientiousness, and Openness scales (Saucier, 1994).—These three personality traits were measured using Saucier's (1994) Big Five mini-markers, which include 40 adjectives tapping the five factors Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness. In the present study, internal consistency reliability and model fit were acceptable: Agreeableness's α was .67, (GFI = .97, AGFI = .92, CFI = .96, RMSEA = .064); Conscientiousness's α was .79, (GFI = .96, AGFI = .90, CFI = .96, RMSEA = .077); Openness's α was .73, (GFI = .97, AGFI = .94, CFI = .99, RMSEA = .027).

Self-control Scale (SCS: Tangney, *et al.*, 2004).—The brief, 13-item version was used to estimate participants' self-control ability. Example items include "I am good at resisting temptation" and "I say inappropriate things" (reverse scored). Participants' higher scores on the scale reflect greater capacity to override their thoughts, feelings, and habitual patterns of behavior. In the present study, Cronbach's α was .75. The one-factor structure fit the present data acceptably: GFI = .90, AGFI = .85, CFI = .88, RMSEA = .064.

Translation

Except for the Chinese version of the Self-Efficacy Scale (Schwarzer, Bäßler, Kwiatek, Schroder, & Zhang, 2008), other scales used in Study 2 were translated and back-translated by the approach described in Study 1.

RESULTS

To assess the factor structure of the GDGQ, confirmatory factor analysis procedures were conducted using structural equation modeling in Sample 1 by using maximum likelihood estimation (Arbuckle & Wothke, 1999). Items were retained based on two indices: the modification index of each item ($MI > 4$) and the factor loading ($\lambda > .30$). The result verified the two-factor structure obtained in the EFA and had acceptable fit: GFI = .95, AGFI = .90, CFI = .89, RMSEA = .079 (Sample 1); GFI = .87, AGFI = .87, CFI = .87, RMSEA = .08 (Sample 2). Moreover, the chi-square test of differences indicated that the two-factor model provided a statistically significantly better fit than the one-factor model: Sample 1, $\Delta x^2 = 41.88$ ($p < .001$); Sample 2, $\Delta x^2 = 61.40$ ($p < .001$). All items loaded statistically significantly ($ps < .001$) on the latent variable. Factor loadings ranged from .30 to .73. Thus, the results of the EFA and CFA supported the hypothesis that delay of gratification had a two-factor structure, Controlling-Impulse and Planning-and-Waiting.

Reliability and Validity

Cronbach's α and McDonald's ω were calculated for each scale (Table 2 and Table 3) (Zinbarg, Yovel, Revelle, & McDonald, 2006). Many scales are assumed to be primarily a measure of one latent variable. If that is true, the latent variable should account for the majority of the variance in the scale scores. Omega is calculated based on confirmatory factor analysis and is a more accurate estimate (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005).

The internal consistency reliabilities of Controlling-Impulse and the overall scale were acceptable (Cronbach's $\alpha = .70$ to $.75$; McDonald's $\omega = .70$ to $.78$; see Tables 2 and 3), while Planning-and-Waiting had relatively poorer internal consistency (Cronbach's $\alpha = .60$; McDonald's $\omega = .62$ to $.64$). The two subscales were weakly related in both samples (Sample 1, $r = .19$; Sample 2, $r = .21$). Item-total correlations were low to moderate, ranging from $.28$ to $.63$ (Sample 1) and $.33$ to $.61$ (Sample 2).

Table 2 and Table 3 display the descriptive statistics and correlations among measures. Controlling-Impulse and Planning-and-Waiting showed different relationships with validity measures. In Sample 1, Uncertainty Avoidance was significantly and positively related with Planning-and-Waiting, but not to Controlling-Impulse. Self-efficacy was significantly and positively related to Planning-and-Waiting and delay of gratification, but not to Controlling-Impulse. In Sample 2, Lnk was significantly and negatively related to Planning-and-Waiting, but not to Controlling-Impulse: with more willingness to wait for the possible larger reward and plan for the future, the delay discounting rate was lower. Agreeableness, Conscientiousness, and Self-control were significantly and positively related to Planning-and-Waiting and Controlling-Impulse, whereas Openness was only significantly and positively related with Planning-and-Waiting, but not with Controlling-Impulse.

CFC-Total and CFC-Future were significantly and positively related with Planning-and-Waiting and Controlling-Impulse, and CFC-Immediate was significantly and negatively related to both the scales (Table 2). A series of multiple regression analyses were run to examine the unique contributions of the Planning-and-Waiting and Controlling-Impulse subscales in predicting CFC-Future and CFC-Immediate. For CFC-Immediate, age and gender were controlled in Step 1; in Step 2, Controlling-Impulse and Planning-and-Waiting were entered together. Results showed that the only unique predictor of CFC-Immediate was Controlling-Impulse ($\beta = -.18$, $t_{319} = -3.55$, $p < .01$), not Planning-and-Waiting ($p = -.09$). In the same way, CFI-Future was found to be predicted by both Planning-and-Waiting ($\beta = .18$, $t_{319} = 4.30$, $p < .01$) and Controlling-Impulse ($\beta = .14$, $t_{319} = 3.59$, $p < .01$).

As for divergent validity, the result indicated that the average variance extracted (AVEs) of the two factors in Sample 1 and 2 were $.28$ and

TABLE 3
DESCRIPTIVE STATISTICS AND PEARSON CORRELATIONS FOR SAMPLE 2 IN STUDY 2 (N = 165)

Variable	M	SD	Correlations									
			1	2	3	4	5	6	7	8	9	10
1. Age	27.07	3.57	—									
2. Gender	1.54	0.50	.08	—								
3. PW	4.75	0.97	.07	-.09	(.60/.64)							
4. CI	4.74	0.95	.12	-.06	.21†	(.70/.70)						
5. Deferment of gratification	4.71	0.75	.11	-.10	.62†	.88†	(.70/.71)					
6. Lnk	-4.00	1.48	.00	.10	-.20*	-.14	-.20*	—				
7. Agreeableness	5.65	0.70	.09	.05	.29†	.25†	.32†	.02	(.67/.75)			
8. Conscientiousness	5.08	0.86	.19	-.12	.37†	.31†	.42†	-.13	.49†	(.79/.79)		
9. Openness	5.03	0.70	-.18+	-.27++	.34†	.14	.28†	-.15	.41†	.44†	(.73/.73)	
10. Self-control	4.74	0.81	.17+	.02	.38†	.40†	.51†	-.11	.24†	.41†	.31†	(.75/.88)

Note.—Gender: men = 1; women = 2; CI = Controlling-Impulse; PW = Planning-and-Waiting; Lnk = Delay-Discounting Rate. Cronbach's alphas and McDonald's omegas (α/ω) are on the diagonal in brackets. * $p < .05$. † $p < .01$.

.38, which was larger than the square of the correlation between the two factors (Sample 1, .16; Sample 2, .04), indicating that divergent validity is acceptable (Formell & Larcker, 1981).

Study 3

In Study 3, the two-factor model's predictive validity was examined via correlation analysis and linear regression. It was hypothesized that these two factors would function differently in predicting employees' long-term achievement in organizations, such as performance and creative performance. In the first two studies, self-report data from the same source was used; here, to reduce common method bias, supervisors' ratings of performance and creative performance were elicited three months after the employees were administered the self-report measures.

METHOD

Participants and Procedure

An independent sample, including 85 participants (49 men, 36 women) were recruited from four private and foreign Chinese information technology companies. The average age was 27.7 yr. ($SD = 3.2$), and 97.6% had a college diploma or higher degree. Participants were classified by job position: 63 were front-line employees (74.1%), 16 were mid-level managers (18.8%), four were senior managers (4.7%) and two did not report their job responsibilities (2.4%).

To reduce common method errors, data were collected in two waves. All employees were asked to complete the GDGQ and the three scales from the Big Five personality scale (Time 1). Three months later (Time 2), their supervisor-rated job performance and creative performance scores were collected.

Measures

Delay of gratification.—The 11-item two-factor GDGQ established in Study 1 and 2 measures delay of gratification.

Job performance.—Farh and Cheng's (1997) 4-item superior-rating performance scale was used as the measure of job performance. Superiors rated their subordinates (e.g., "He/she is one of the best employees in my department") on a 7-point Likert-type scale with anchors 1: Very strongly disagree and 7: Very strongly agree. The translated Chinese version had been used in the past (e.g., Aryee & Chen, 2006). In the present study, Cronbach's α was .89.

Creative performance.—Creative performance was assessed by George and Zhou's (2001) 13-item scale. Superiors rated their subordinates (e.g., "Suggests new ways to increase quality") on a 7-point Likert-type scale with anchors 1: Very strongly disagree and 7: Very strongly agree. The

translated Chinese version has been used in the past (e.g., Zhou, Shin, Brass, Choi, & Zhang, 2009; Wang & Cheng, 2010). In the present study, Cronbach's α was .89.

RESULTS

Table 4 displays descriptive statistics and correlations among measures. Job performance and creative performance were significantly and positively correlated to Controlling-Impulse and delay of gratification, but were not statistically significantly correlated to Planning-and-Waiting. Hierarchical regression analyses were conducted to assess the function of Controlling-Impulse and Planning-and-Waiting in predicting job performance and creative performance. Both the Controlling-Impulse and Planning-and-Waiting scores were mean-centered.

Predicting job performance ratings, Step 1 included control variables (age was deleted due to multicollinearity, $VIF > 10$) (Neter, Wasserman, & Kutner, 1990). Step 1 did not reached statistical significance ($R^2 = .08$, $p > .05$), but job position was a statistically significant predictor of job performance ratings ($\beta = .36$, $p < .05$). In Step 2, performance was regressed on Controlling-Impulse and Planning-and-Waiting. This step produced statistically significant results ($R^2 = .20$, $p < .01$; $\Delta R^2 = .12$, $\Delta F_{2,79} = 8.28$, $p < .01$). Controlling-Impulse was a statistically significant predictor of job performance ($\beta = .30$, $p < .01$).

Similarly, a regression was conducted to predict creative performance ratings. Step 1 was not statistically significant ($R^2 = .07$, $p > .10$), but Step 2 was ($R^2 = .25$, $p < .01$; $\Delta R^2 = .18$, $\Delta F_{2,79} = 12.28$, $p < .01$). Again, Controlling-Impulse scores were a statistically significant predictor of creative performance ratings ($\beta = .32$, $p < .01$). Taken together, Controlling-Impulse, but not Planning-and-Waiting, predicted both job performance and creative performance ratings, thus supporting a differentiation of the two factors of delay of gratification.

GENERAL DISCUSSION

The GDGQ is meant to measure delay of gratification among adolescents and adults, and has been regarded as a unidimensional scale. The purpose of the present study was to explore the possibility of a multidimensional structure. Data from four separate samples, with 894 participants in total, were obtained with multi-wave and multi-source methods. The results showed that the two-factor model fit the data better than the single-factor model. The two factors, labeled Controlling-Impulse and Planning-and-Waiting, had different correlation patterns with Openness, delay discounting, self-efficacy and uncertainty avoidance. Meanwhile, Controlling-Impulse predicted statistically significant amounts of variance in performance and creative performance, whereas Planning-and-

TABLE 4
DESCRIPTIVE STATISTICS AND PEARSON'S CORRELATIONS IN STUDY 3 ($N = 85$)

Variable	M	SD	Correlations							
			1	2	3	4	5	6	7	8
1. Age	27.73	3.23								
2. Gender	1.42	0.50	-.20							
3. Job position	1.26	0.58	.47†	-.10						
4. Job performance	4.70	0.85	.16	-.07	.23*	(.89/.85)				
5. Creative performance	4.30	0.73	.12	-.03	.26*	.74†	(.89/.88)			
6. PW	5.32	0.67	.32†	-.32†	.19	.03	.07	(.60/.65)		
7. CI	4.19	0.90	-.35†	.16	-.03	.31†	.41†	-.09	(.71/.70)	
8. DG	4.75	0.54	-.09	.06	.09	.28†	.39†	.55†	.78†	(.70/.79)

Note.—Gender: men = 1; women = 2; CI = Controlling-Impulse; PW = Planning-and-Waiting; DG = Deferment of Gratification. Cronbach's alphas and McDonald's omegas (α/ω) are on the diagonal in brackets. * $p < .05$. † $p < .01$.

Waiting did not. These results are consistent with recent arguments that similar constructs such as time orientation also have multiple dimensions (e.g., Zimbardo & Boyd, 1999; Joireman, *et al.*, 2008; Joireman, *et al.*, 2012).

With respect to evidence of convergent validity, the present research showed that the subscales Planning-and-Waiting and Controlling-Impulse both had statistically significant correlations with a measure of self-control, indicating that both of these two components are involved in the self-regulation process. Dramatically, although delay discounting is often used as an equivalent measure of individuals' ability to delay gratification (Reynolds, *et al.*, 2002), the understanding of the precise correspondence between delay of gratification and delay discounting is limited (Kirby, *et al.*, 2005). The findings may explain why the relationship between discounting rate and ability to delay gratification is mixed: because only one component of delay of gratification, Planning-and-Waiting, was significantly correlated with delay discounting rate. This finding is consistent with previous claims that delay discounting is more related to future planning (Schweizer, 2002; Scholten & Read, 2006; Zauberman, Kim, Malkoc, & Bettman, 2009) rather than impulse control (Steinberg, *et al.*, 2009).

As for Consideration of Future Consequences, each subscale was statistically significantly correlated with the two components of GDGQ. The two GDGQ subscales were differentially predictive of the two CFC subscales. Planning-and-Waiting only predicted statistically significant variance in CFC-Immediate (positively), in line with the hypothesis. In contrast, Controlling-Impulse predicted CFC-Immediate (negatively) and CFC-Future (positively). These results can be explained using the two-phase model of delay of gratification (Mischel, 1974). During phase 1, controlling an impulse, "delayers" would make the choice to abandon immediate gratification for the sake of delayed but more valuable outcome, ignoring the importance of the immediate consequences of behavior (low CFC-Immediate) and attaching importance to the future consequences of behavior (high CFC-Future). During phase 2, Planning-and-Waiting, "delayers" maintain their choice until the eventual goal is achieved (Mischel, 1974). People must focus on the future consequence of their choices to help them better plan and achieve long-term goals (high CFC-Future). Although the mechanism should be further examined, the different relationship between CFC subscales and delay of gratification dimensions provides supportive evidence for the two-factor structure.

In terms of criterion-related validity, the present study provided strong evidence that the subscales correlate with psychological variables in theoretically predictable ways. Uncertainty avoidance and self-efficacy were statistically significantly correlated with Planning-and-Waiting but not with Controlling-Impulse, which provides further evidence for the

two-factor structure of delay of gratification. As for three of the Big Five personality traits, only the measure of Openness showed discriminative correlation with the two factors of delay of gratification. Previous research has supposed that participants with Openness can manage to delay gratification because they can avoid focusing on the possibility of an immediate reward (Krueger, *et al.*, 1996). However, the results of the present study refuted this view and clarified that participants with Openness could delay gratification by planning and waiting for a large, delayed reward rather than by controlling impulses on an immediate reward.

In addition, this study advances understanding of the functions of delay of gratification. Although Mischel (1974) proposed a two-phase model for delay of gratification, there is no research explaining which phase is more important in the process of delay of gratification. Compared with Planning-and-Waiting, Controlling-Impulse was a stronger predictor of supervisor's ratings of employees' performance and creative performance. Mischel's view of emphasizing the important role of Controlling-Impulse on task performance (Mischel, 1983) and the findings that impulsivity impairs performance in completing reasoning tasks (Schweizer, 2002), offer parallel explanations for this result. For objective reasons (e.g., chances of promotion) and subjective reasons (e.g., whether one's superior is favorable or not) in an organization, people with high delay of gratification are not guaranteed to get what they wait and plan for in their careers (Pogson, Cober, Doverspike, & Rogers, 2003). Therefore, just waiting and planning cannot itself lead to good job performance and creative performance.

Limitations and Future Directions

The study has several limitations. Firstly, the subscale Planning-and-Waiting had poor internal consistency reliability (Cronbach's $\alpha = .60$); McDonald's ω was also poor (.62 to .64). However, Planning-and-Waiting rather than Controlling-Impulse had statistically significant correlations with uncertainty avoidance, delay discounting rate, self-efficacy, and Openness. Therefore, Planning-and-Waiting is an indispensable factor of delay of gratification. It differentiates general delay of gratification from impulsivity. The low reliability may be due to the small number of items, so future research could add new items to increase the scale's internal consistency. Secondly, the study was conducted in Chinese samples characterized by cultural features like long-term orientation (Hofstede, 2007) and abstinence. Thus, the generality of the two-factor structure of GDGQ must be verified in other cultures. Finally, some items of Controlling-Impulse are related to spending money, such as "I like to spend my money as soon as I get it." Ray and Najman (1986) developed the GDGQ including sev-

eral items describing spending because they believed that "those who invest in their futures by saving their financial resources are more likely to be achievers and to be successful." Being able to control the impulse to spend money is an important expression of delay of gratification. While there are other domains which could be included in delay of gratification such as food, physical pleasures, social interactions, achievement, and so on (Hoerger, Quirk, & Weed, 2011), the factor was named Controlling-Impulse rather than Controlling-Spending for two reasons: (1) The goal was to assess the two-dimensional structure of the delay of gratification scale corresponding to Mischel's (1974) two-phase model, in which the first phase is to control the impulse toward an immediate reward and the second phase is to plan and wait for a future reward. Thus, a more general impulse was represented by the intention in the first stage of delay of gratification. (2) The items of Controlling-Impulse cover deferment of gratification in purchases as well as in other areas (e.g., "Would you describe yourself as often being too impulsive for your own good?" (Norvilitis & MacLean, 2010). Impulsive spending and buying are not equivalent. Future studies should extend the content of Controlling-Impulse to other related domains.

Conclusion

In conclusion, based on exploratory and confirmatory factor analysis, the present study provided evidence that delay of gratification has a two-dimension structure, described as Controlling-Impulse and Planning-and-Waiting. These two subscales had distinguishable correlations with delay-discounting rate, Openness, uncertainty avoidance, two Consideration of Future Consequences subscales, and self-efficacy. Furthermore, Controlling-Impulse and Planning-and-Waiting functioned differently in predicting job performance and creative performance as rated by superiors.

REFERENCES

- ARBUCKLE, J., & WOTHKE, W. (1999) *Amos 4.0 user's guide*. Chicago, IL: SmallWaters.
- ARYEE, S., & CHEN, Z. X. (2006) Leader-member exchange in a Chinese context: antecedents, the mediating role of psychological empowerment and outcomes. *Journal of Business Research*, 59, 793-801.
- AYDUK, O. (2007) Delay of gratification in children. In Y. Shoda, D. Cervone, & G. Downey (Eds.), *Persons in context: building a science of the individual*. New York: Guilford Press. Pp. 97-109.
- BANDURA, A. (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- BEMBENUTTY, H., & KARABENICK, S. (1998) Academic delay of gratification. *Learning and Individual Differences*, 10, 329-346.
- BRACKEN, B. A. T., & BARONA, A. (1991) State of the art procedures for translating, validating, and using psycho-educational tests in cross-cultural assessment. *School Psychology International*, 2, 119-132.

- CALDWELL, L. M., & MOWRER, R. R. (1998) The link between procrastination, delay of gratification, and life satisfaction: a preliminary analysis. *Psi Chi Journal of Undergraduate Research*, 3, 145-150.
- DECOSTER, J. (1998) *Overview of factor analysis*. University of Alabama, Department of Psychology Retrieved from: <http://www.stat-help.com/notes.html>.
- DORFMAN, P., & HOWELL, J. (1988) Dimensions of national culture and effective leadership patterns: Hofstede revisited. *Advances in International Comparative Management*, 3, 127-150.
- DUCKWORTH, A. L., & SELIGMAN, M. E. P. (2005) Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16, 939-944.
- FARH, J., & CHENG, B. (1997) Modesty bias in self-ratings in Taiwan: impact of item wording, modesty value, and self-esteem. *Chinese Journal of Psychology*, 39, 103-118.
- FORNELL, C., & LARCKER, D. F. (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- GEORGE, J., & ZHOU, J. (2001) When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of Applied Psychology*, 86, 513-524.
- HINKIN, T. (1998) A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1, 104-121.
- HOERGER, M., QUIRK, S. W., & WEED, N. C. (2011) Development and validation of the Delaying Gratification Inventory. *Psychological Assessment*, 23, 725-738.
- HOFSTEDE, G. (1991) *Organizations and cultures: software of the mind*. New York: McGraw-Hill.
- HOFSTEDE, G. (2007) A European in Asia. *Asian Journal of Social Psychology*, 10, 16-21.
- JERUSALEM, M., & SCHWARZER, R. (1992) *Self-efficacy: thought control of action*. Washington, DC: Hemisphere.
- JOIREMAN, J., BALLIET, D., SPROTT, D., SPANGENBERG, E., & SCHULTZ, J. (2008) Consideration of future consequences, ego-depletion, and self-control: support for distinguishing between CFC-Immediate and CFC-Future sub-scales. *Personality and Individual Differences*, 45, 15-21.
- KAISER, H. F. (1974) An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- KERPELMAN, J. L., & MOSHER, L. S. (2004) Rural African American adolescents' future orientation: the importance of self-efficacy, control and responsibility, and identity development. *An International Journal of Theory and Research*, 4, 187-208.
- KIRBY, K. N., & MARAKOVI, N. N. (1996) Delay-discounting probabilistic rewards: rates decrease as amounts increase. *Psychonomic Bulletin & Review*, 3, 100-104.
- KIRBY, K. N., PETRY, N. M., & BICKEL, W. K. (1999) Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology*, 128, 78-87.
- KIRBY, K. N., WINSTON, G. C., & SANTISTEBAN, M. (2005) Impatience and grades: delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences*, 15, 213-222.
- KRUEGER, R., CASPI, A., MOFFITT, T., WHITE, J., & STOUTHAMER-LOEBER, M. (1996) Delay of gratification, psychopathology, and personality: is low self-control specific to externalizing problems? *Journal of Personality*, 64, 107-129.
- LOK, M. H., BOND, A. J., & TSE, W. S. (2009) Contrasting effects of a hot and a cool system in anger regulation on cooperative behaviours. *International Journal of Psychology*, 44, 333-341.

- McDONALD, R. P. (1999) *Test theory: a unified treatment*. Hillsdale, NJ: Erlbaum.
- METCALFE, J., & MISCHEL, W. (1999) A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, 106, 3-19.
- MILLER, M., WOEHR, D., & HUDSPETH, N. (2002) The meaning and measurement of work ethic: construction and initial validation of a multidimensional inventory. *Journal of Vocational Behavior*, 60, 451-489.
- MISCHEL, W. (1961) Father-absence and delay of gratification: cross-cultural comparisons. *The Journal of Abnormal and Social Psychology*, 63, 116-124.
- MISCHEL, W. (1974) Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 7. New York: Academic Press. Pp. 249-292.
- MISCHEL, W. (1981) Metacognition and the rules of delay. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: frontiers and possible futures*. New York: Cambridge Univer. Press. Pp. 240-271.
- MISCHEL, W., & MISCHEL, H. N. (1983) The development of children's knowledge of self-control strategies. *Child Development*, 54, 603-619.
- MISCHEL, W., SHODA, Y., & PEAKE, P. (1988) The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, 54, 687-696.
- MISCHEL, W., SHODA, Y., & RODRIGUEZ, M. (1989) Delay of gratification in children. *Science*, 244, 933-938.
- MURAVEN, M., & BAUMEISTER, R. F. (2000) Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin*, 126, 247-259.
- NETER, J., WASSERMAN, W., & KUTNER, M. (1990) *Applied linear models*. Homewood, IL: Irwin.
- NORVILITIS, J. M., & MACLEAN, M. G. (2010) The role of parents in college students' financial behaviors and attitudes. *Journal of Economic Psychology*, 31, 55-63.
- OLDHAM, G., & CUMMINGS, A. (1996) Employee creativity: personal and contextual factors at work. *Academy of Management Journal*, 39, 607-634.
- PARKE, A., GRIFFITHS, M., & IRWING, P. (2004) Personality traits in pathological gambling: sensation seeking, deferment of gratification and competitiveness as risk factors. *Addiction Research & Theory*, 12, 201-212.
- PETROCELLI, J. V. (2003) Factor validation of the Consideration of Future Consequences Scale. *Journal of Social Psychology*, 143, 404-413.
- POGSON, C., COBER, A., DOVERSPIKE, D., & ROGERS, J. (2003) Differences in self-reported work ethic across three career stages. *Journal of Vocational Behavior*, 62, 189-201.
- RACHLIN, H. (2000) *The science of self-control*. Cambridge, MA: Harvard Univer. Press.
- RACHLIN, H., BROWN, J., & CROSS, D. (2000) Discounting in judgments of delay and probability. *Journal of Behavioral Decision Making*, 13, 145-159.
- RAUCH, A., FRESE, M., & SONNENTAG, S. (2000) Cultural differences in planning-success relationships: a comparison of small enterprises in Ireland, West Germany, and East Germany. *Journal of Small Business Management*, 38, 28-41.
- RAY, J., & NAJMAN, J. (1986) The generalizability of deferment of gratification. *Journal of Social Psychology*, 126, 117-119.
- REYNOLDS, B., DE WIT, H., & RICHARDS, J. B. (2002) Delay of gratification and delay discounting in rats. *Behavioural Processes*, 59, 157-168.
- ROSENBAUM, M., & BEN-ARI SMIRA, K. (1986) Cognitive and personality factors in the delay of gratification of hemodialysis patients. *Journal of Personality and Social Psychology*, 51, 357-364.

- SAUCIER, G. (1994) Mini-markers: a brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506-516.
- SCHOLTEN, M., & READ, D. (2006) Discounting by intervals: a generalized model of intertemporal choice. *Management Science*, 52, 1424-1436.
- SCHWARZER, R., BÄBLER, J., KWIATEK, P., SCHRODER, K., & ZHANG, J. X. (2008) The assessment of optimistic self-beliefs: comparison of the German, Spanish, and Chinese versions of the General Self-efficacy Scale. *Applied Psychology*, 46, 69-88.
- SCHWEIZER, K. (2002) Does impulsivity influence performance in reasoning? *Personality and Individual Differences*, 33, 1031-1043.
- SHALLEY, C. E., & ZHOU, J. (2008) Organizational creativity research: a historical overview. In J. Zhou & C. E. Shalley (Eds.), *Handbook of organizational creativity*. London, UK: Lawrence Erlbaum Associates. Pp. 3-31.
- STEINBERG, L., GRAHAM, S., O'BRIEN, L., WOOLARD, J., CAUFFMAN, E., & BANICH, M. (2009) Age differences in future orientation and delay discounting. *Child Development*, 80, 28-44.
- STRATHMAN, A., GLEICHER, F., BONINGER, D. S., & EDWARDS, C. S. (1994) The consideration of future consequences: weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, 66, 742-752.
- TANGNEY, J. P., BAUMEISTER, R. F., & BOONE, A. L. (2004) High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271-324.
- WANG, A. C., & CHENG, B. S. (2010) When does benevolent leadership lead to creativity? The moderating role of creative role identity and job autonomy. *Journal of Organizational Behavior*, 31, 106-121.
- WARD, W., PERRY, T., WOLTZ, J., & DOOLIN, E. (1989) Delay of gratification among black college student leaders. *Journal of Black Psychology*, 15, 111-128.
- WITT, L. A. (1990a) Delay of gratification and locus of control as predictors of organizational satisfaction and commitment: sex differences. *The Journal of General Psychology*, 117, 437-446.
- WITT, L. A. (1990b) Person-situation effects and gender differences in the prediction of social responsibility. *Journal of Social Psychology*, 130, 543-553.
- ZAUBERMAN, G., KIM, B. K., MALKOC, S. A., & BETTMAN, J. R. (2009) Discounting time and time discounting: subjective time perception and intertemporal preferences. *Journal of Marketing Research*, 46, 543-556.
- ZHOU, J., SHIN, S. J., BRASS, D. J., CHOI, J., & ZHANG, Z. X. (2009) Discounting time and time discounting: subjective time perception and intertemporal preferences. *Journal of Applied Psychology*, 94, 1544-1552.
- ZIMBARDO, P. G., & BOYD, J. N. (1999) Putting time in perspective: a valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, 77, 1271-1288.
- ZINBARG, R. E., REVELLE, W., YOVEL, I., & LI, W. (2005) Cronbach's α , Revelle's β , and McDonald's ω_h : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123-133.
- ZINBARG, R. E., YOVEL, I., REVELLE, W., & McDONALD, R. P. (2006) Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ω_h . *Applied Psychological Measurement*, 30, 121-144.

GENDER INVARIANCE OF THE COLLEGE STUDENT STRESS SCALE¹

RONALD C. FELDT AND CHRISTINA UPDEGRAFF

Mount Mercy University

Summary.—Assessment of perceived stress may be an important prerequisite to deployment of effective coping in efforts to help college students adjust to academic and social demands of college. The study examined the extent to which a seven-item measure of the College Student Stress Scale is invariant across gender. Results indicated invariance of factor loadings, factor variance, and all but one item intercept. No statistically significant gender difference was observed between latent variable means.

The inability to manage the stress response in college is associated with poor college adjustment, depression, and reduced life satisfaction (Segrin, Hanzal, Donnerstein, Taylor, & Domschke, 2007; Asberg, Bowers, Renk, & McKinney, 2008; Verschoor & Markus, 2011; Credé & Niehorster, 2012). Assessment of stress is an important component of efforts to assist students in benefiting from the college experience, and instruments used for such assessment should demonstrate reliability and construct validity. The College Student Stress Scale was developed to provide a brief and global measure of college stress that includes appraisal of the ability to maintain control (Feldt, 2008). A recent study confirmed that the College Student Stress Scale has a two-factor structure; however, use of a three-item subscale for the second factor was not recommended due to low reliability (Feldt & Koch, 2011). Structural and external aspects of construct validity (Messick, 1995) were established in two previous studies (Feldt, 2008; Feldt & Koch, 2011). One question that remains is the extent to which the College Student Stress Scale is invariant across gender, the generalizability aspect of construct validity (Messick, 1995). Invariance testing at construct and item levels provides tests of generalizability across populations, and multigroup confirmatory factor analysis (CFA) is commonly used for such testing (Dimitrov, 2010; Sass, 2011). Comparison of gender means based on observed scores assumes measurement invariance, and results of such comparisons when the assumption is violated may be erroneous (Sass, 2011). The purpose of the present study was to examine measurement invariance across gender of the seven-item subscale.

¹Address correspondence to Ronald C. Feldt, Department of Psychology, Social Work, Sociology, and International Studies, Mount Mercy University, Cedar Rapids, IA 52402 or e-mail (rfeldt@mtmercy.edu).

Method

Participants included 460 undergraduate college students (353 women and 107 men; 365 first-year students and more than 92% Euro-American) who were enrolled in psychology classes. Age ranged from 17 to 44 years ($M = 19.5$ yr.; $SD = 3.3$). No gender difference in age was observed ($p > .05$). The College Student Stress Scale has 11 items and a Likert-type scale with response categories ranging from 1: *Never* to 5: *Very often*. It was administered in a classroom setting.

Results and Discussion

Data for this analysis were combined from two previous studies (Feldt, 2008; Feldt & Koch, 2011) to create a large enough sample to support an invariance analysis. Multigroup confirmatory factor analysis was used with maximum likelihood estimation. A principal component analysis was used to identify a reference indicator based on the most similar component loadings across gender (Finch & French, 2008). This study created a well-fitting model and determined goodness of fit for each gender, in addition to a combined configural model. A correlated error for Items 7 and 8 was estimated (Feldt & Koch, 2011). Fit indexes included the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). The chi-squared test was not used to assess goodness of fit because of its tendency to be influenced by sample size and to reject hypothesized models when negligible differences exist (Brown, 2006; West, Taylor, & Wu, 2012). A series of constrained models was created with equality constraints on factor loadings, item intercepts, and factor variance. Change in CFI ($< -.01$) (Cheung & Rensvold, 2002) was used for model comparison because it is less likely than the chi-squared difference test ($\Delta\chi^2$) to result in Type I errors when models do not have perfect fit (French & Finch, 2011).

Good fit was indicated by CFI values $> .95$ and SRMR values $< .05$ for the configural, baseline model for each gender and the combined model (Table 1). Standardized regression weights ranged from .42 to .80 for women and from .58 to .86 for men. Results of a series of comparisons of constrained models indicated invariance of the correlated error estimate, factor loadings, and factor variance. In contrast, the intercept for Item 10 was observed to be noninvariant ($\Delta CFI = -.018$), thus indicating differential item functioning (DIF). Improved fit was observed when the intercept was freely estimated. Item 10 intercept estimates for women and men were 1.79 and 2.08, respectively. This indicates that men are more likely to endorse this item ("no longer in control"). All other intercepts were noninvariant. The difference between gender latent means was not statistically significant whether the DIF item was included or was excluded from the analysis,

TABLE 1
MEASUREMENT INVARIANCE OF THE COLLEGE STUDENT STRESS SCALE ACROSS
GENDER ($N = 460$)

Model	χ^2	df	CFI	RMSEA (90%CI)	SRMR
Women	51.18	13	.963	.091 (.066, .118)	.037
Men	20.67	13	.982	.075 (.000, .132)	.034
Combined	71.88	26	.968	.062 (.045, .079)	.037
Factor loadings equal	80.41	32	.966	.057 (.042, .073)	.040
Item intercepts equal	112.99	38	.948	.066 (.042, .073)	.040
Item 10 intercept estimated	100.94	37	.956	.061 (.052, .080)	.040
Factor variance equal	103.40	38	.954	.061 (.047, .076)	.044

Note.— χ^2 = chi-squared goodness-of-fit test, CFI = comparative fit index, RMSEA (with 90% confidence interval) = root mean square error of approximation, and SRMR = standardized root mean square residual.

with estimates of the latent variable mean difference of .15 and .22, respectively (both $p > .05$). In addition, comparison of the fully- (all intercepts constrained) and partially-invariant (intercept for Item 10 was estimated) models indicated a slight increase in the estimated difference to .22 in the partially-invariant model. However, both estimates were not statistically significant ($p > .05$). Our results support the partial measurement invariance of the seven-item measure of college stress with minimal impact of the noninvariant intercept for Item 10 (Byrne, Shavelson, & Muthén, 1989).

The seven-item measure should serve as a global measure of college stress for researchers who wish to investigate relationships between global stress and other constructs. The measure should be particularly useful when researchers employ relatively lengthy instruments within a brief session. Although the second sample (Feldt & Koch, 2011) was slightly older than the first sample (Feldt, 2008) ($p < .001$), parameter estimates for the correlated error, factor loadings, intercepts, and variance of the latent variable were observed to be invariant across the two samples. In contrast, the estimated difference in latent means was greater for the older sample ($p < .05$).

The major limitation includes limited generalizability due to the fact that the sample was comprised primarily of Euro-American women in their first-year of college and Euro-American men. In addition, the sample consisted of students committed to service-oriented majors (nursing, psychology, and social work). Subsequent research should include additional populations that vary in race and ethnicity, in addition to year in college and college major.

REFERENCES

- ASBERG, K. K., BOWERS, C., RENK, K., & MCKINNEY, C. (2008) A structural equation modeling approach to the study of stress and psychological adjustment in emerging adults. *Child Psychiatry and Human Development*, 39, 481-501. DOI: 10.1007/s10578-008-0102-0
- BROWN, T. A. (2006) *Confirmatory factor analysis for applied research*. New York: Guilford.
- BYRNE, B. M., SHAVELSON, R. J., & MUTHÉN, B. (1989) Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466. DOI: 10.1037/0033-2909.105.3.456
- CHEUNG, G. W., & RENSVOOLD, R. B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255. DOI: 10.1207/S15328007SEM0902_5
- CREDÉ, M., & NIEHORSTER, S. (2012) Adjustment to college as measured by the Student Adaptation to College Questionnaire: a quantitative review of its structure and relationships with correlates and consequences. *Educational Psychology Review*, 24, 133-165. DOI: 10.1007/s10648-011-9184-5
- DIMITROV, D. M. (2010) Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43, 121-149. DOI: 10.1177/0748175610373459
- FELDT, R. C. (2008) Development of a brief measure of college stress: the College Student Stress Scale. *Psychological Reports*, 102, 855-860. DOI: 10.2466/pr0.102.3.855-860
- FELDT, R. C., & KOCH, C. (2011) Reliability and construct validity of the College Student Stress Scale. *Psychological Reports*, 108, 660-666. DOI: 10.2466/02.08.13.16.PR0.108.2.660-666
- FINCH, W. H., & FRENCH, B. F. (2008) Using exploratory factor analysis for locating invariant referents in factor invariance studies. *Journal of Modern Applied Statistical Methods*, 7, 223-233. Retrieved at <http://www.jmasm.com/>
- FRENCH, B. F., & FINCH, W. (2011) Model misspecification and invariance testing using confirmatory factor analytic procedures. *Journal of Experimental Education*, 79, 404-428. DOI: 10.1080/00220973.2010.517811
- MESSICK, S. (1995) Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749. DOI: 10.1037/0003-066X.50.9.741
- SASS, D. A. (2011) Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29, 347-363. DOI: 10.1177/0734282911406661
- SEGRIN, C., HANZAL, A., DONNERSTEIN, C., TAYLOR, M., & DOMSCHKE, T. J. (2007) Social skills, psychological well-being, and the mediating role of perceived stress. *Anxiety, Stress & Coping: An International Journal*, 20, 321-329. DOI: 10.1080/10615800701282252
- VERSCHOOR, E., & MARKUS, C. (2011) Effects of acute psychosocial stress exposure on endocrine and affective reactivity in college students differing in the 5-HTTLPR genotype and trait neuroticism. *Stress: The International Journal on the Biology of Stress*, 14, 407-419. DOI: 10.3109/10253890.2010.548886
- WEST, S. G., TAYLOR, A. B., & WU, W. (2012) Model fit and model selection in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford. Pp. 209-231.

TEACHER STRESS QUESTIONNAIRE: VALIDITY AND RELIABILITY STUDY IN ITALY¹

MARIA CLELIA ZURLO, DANIELA PES, AND ROBERTO CAPASSO

*Department of Humanistic Studies
University of Naples "Federico II"*

Summary.—This study analyses the psychometric properties of the Italian version of the Teacher Stress Questionnaire elaborated in England by Travers and Cooper in 1996. This Italian survey was completed by 863 teachers randomly drawn from a cross-section of Italian school levels. The construct validity of the questionnaire was verified by factor analysis and by measuring the internal consistency of the single scales. All dimensions measured by the Teacher Stress Questionnaire were compared for sample subgroups of all teacher levels. Several meaningful and reliable factors emerged from the factor analysis of the scales. The internal consistency of each scale (Cronbach's α) revealed satisfactory values. Teachers' age and school level were determining factors for all dimensions of stress explored. The Italian version of the Teacher Stress Questionnaire showed satisfactory psychometric properties and constitutes a useful and reliable measure to analyse stress in Italian schools.

For several decades and especially in recent years, many studies conducted throughout the world have highlighted that teaching is to be considered a highly stressful occupation (Farber, 1991; Chan & Hui, 1995; Travers & Cooper, 1996; Benmansour, 1998; Dunham & Varma, 1998; Guglielmi & Tatrow, 1998; Pithers & Soden, 1998; Byrne, 1999; Kyriacou, 2001; Van Dick & Wagner, 2001; Bauer, Stamm, Virnich, Wissing, Kriston, Müller, *et al.*, 2006; Unterbrink, Hack, Pfeifer, Buhl-Grießhaber, Müller, Wesche, *et al.*, 2007; Zurlo, Pes, & Cooper, 2007; Zurlo, Pes, & Siegrist, 2010; Zurlo & Pes, 2012). According to Unterbrink, *et al.* (2007), 22% of German teachers found their occupation extremely stressful. Smith, Brice, Collins, Matthews, and McNamara (2000) found that 41.5% of English teachers report high occupational stress, compared to 32% of nurses, 28% of managers, and 27% of administrators. In Italy, it was found that almost 25% of teachers perceived high stress (Zurlo, 2013), and the percentage was 70% among temporary teachers (Zurlo & Pes, 2012).

In the panorama of occupational stress models, several studies have explored the mechanisms for regulating the interactions between the person and his/her job-environment, which modulate individuals' perceptions of stress (Lazarus, 1999; Dewe, Leiter, & Cox, 2000). With particular reference to stress in teaching, most research has identified many situa-

¹Address correspondence to Maria Clelia Zurlo, Department of Humanistic Studies, University of Naples "Federico II," Via Porta di Massa 1, 80133, Naples, Italy or e-mail (zurlo@unina.it).

tional factors perceived as sources of stress, such as role ambiguity, work overload connected to the management of pupil behaviour and special educational needs, poor school climate, lack of decision latitude, and lack of support. In the Italian school context, further sources of stress are linked to the changes produced by educational reforms promulgated in recent years, requiring teachers to assimilate proposed changes in the curriculum and to gain new expertise, and introducing appraisal of teaching. Moreover, several personal dispositional factors have been identified which influence the cognitive appraisal of stressful events: self-esteem, locus of control, coping strategies, and Type A behaviour (Travers & Cooper, 1996).

From this perspective, the model of occupational stress elaborated by Sutherland and Cooper (1988) constitutes a relevant theoretical and descriptive effort. This model of occupational stress is consistent with an interactive perspective and is also the basis for the specific model of stress in teaching elaborated by Travers and Cooper (1996). The underlying premise of this approach to teacher stress is that undesirable responses (e.g., psychological strain, alcohol consumption, and intention to leave the profession) to the pressure in the working environment (the teaching profession) result from a lack of fit between the dispositional characteristics of the individual teacher and the situational aspects of the job that he or she performs. According to Travers and Cooper's approach, stress is considered a dynamic relational concept, which depends on the constant interplay among different factors that influence each other: (1) individual facets, e.g., age, gender, Type A personality (Friedman & Rosenman, 1959); (2) situational facets, e.g., level of students taught,² age of teaching, number of pupils taught, number of hours worked, sources of pressure in teaching; (3) factors which moderate the interaction between individual and situational facets, e.g., perceived job satisfaction and coping strategies; and (4) effects of teacher stress, e.g., psychological strain (anxiety and depression) and teachers' health-related behavioural responses to stress (smoking, drinking, and intention to leave the profession).

The Teacher Stress Questionnaire

To operationalize their interactive model of teacher stress and to measure the mutual influence existing between stressful situations and individuals' responses to those situations, Travers and Cooper (1996) created the Teacher Stress Questionnaire. This questionnaire comprises six sections. Section 1 includes measures of the respondent's personal and job demographics and biographical items (e.g., sex and age), professional and career information (e.g., years in teaching, school level), school (e.g., number of pupils) and job details (e.g., number of hours worked). Additional

²Levels: nursery school, primary school, junior high school, and high school.

measures are also obtained on health-related behaviours (e.g., smoking, drinking) and on intention to leave the profession. Section 2 is a measure of perceived mental ill-health, the Crown-Crisp Experiential Index (Crown & Crisp, 1979). Section 3 is the Type A Behavioural Style Inventory (Bortner, 1969). Section 4 is the Job Satisfaction Scale (Warr, Cook, & Wall, 1979). Section 5 is the Sources of Pressure in Teaching Scale (Travers & Cooper, 1996). Section 6 is the Coping Style Inventory (Cooper, Sloan, & Williams, 1988).

The Crown-Crisp Experiential Index (formerly the Middlesex Hospital Questionnaire) (Crown & Crisp, 1979) measures psychological well-being and mental health. This inventory is composed of six subscales measuring Free-floating anxiety (8 items), Somatic anxiety (8 items), Depression (8 items), Phobic anxiety (8 items), Obsessionality (8 items), and Hysteria (8 items). Within the Teacher Stress Questionnaire only the first three scales are used, plus a total Overall mental ill-health score, i.e., a total of 24 items' ratings. For this study, reliability values (Cronbach's α) were: Overall mental ill-health, $\alpha = .87$; Free-floating Anxiety, $\alpha = .73$; Somatic Anxiety, $\alpha = .66$; Depression, $\alpha = .70$.

The Bortner's Type A Behavioural Style Inventory (Bortner, 1969) is employed as a measure of behavioural style. The scale consists of 14 bipolar adjectival items measured on an 11-point Likert-type rating scale. The inventory yields a single score ranging from 14 to 154 (low to high Type A behaviour). The presence of this scale within the Teacher Stress Questionnaire is consistent with the underlying premise made by Travers and Cooper that Type A behaviour influences teacher stress.

The Job Satisfaction Scale (Warr, *et al.*, 1979) is used to measure job satisfaction. This scale consists of 15 items, measured on a seven-point Likert-type rating scale for each item, assessing job satisfaction with anchors 1: Extremely dissatisfied and 7: Extremely satisfied. As stated above, the model elaborated by Travers and Cooper considers job satisfaction to be one of the factors that moderates the stress stemming from the interaction between individual facets and situational facets.

The Sources of Pressure in Teaching Scale (Travers & Cooper, 1996) was adopted to explore the particular kinds of job pressure experienced by teachers, i.e., the situational features which influence teacher stress. This scale consists of 98 items describing sources of pressure which, on the one hand, were drawn from 40 transcribed semi-structured interviews conducted in seven British schools, and, on the other hand, have been identified as those factors most reported by teachers in international studies. The items are measured on a six-point Likert-type scale with anchors 1: Strongly disagree is a source of pressure and 6: Strongly agree is a source of pressure.

The Coping Style Inventory is employed to measure teachers' strategies for coping with job stress and derives from the Occupational Stress Indicator (Cooper, *et al.*, 1988). The scale consists of 28 items rated on a six-point Likert-type scale with anchors 1: Never used by me and 6: Very extensively used by me. As above, coping style according to the model elaborated by Travers and Cooper is considered to be a second factor which moderates the stress caused by the interaction between individual and situational facets.

Goals of This Study

The present study presents the Italian version of the Teacher Stress Questionnaire (Travers & Cooper, 1996), and analyses both its psychometric characteristics and factorial structure. In particular, the purpose was firstly to analyse the Teacher Stress Questionnaire's internal consistency reliability by means of item analysis and by measuring the internal consistency of its single scales. Secondly, the construct validity of the questionnaire would be assessed by means of a factor analysis of its single scales. Thirdly, the differences emerging in the dimensions evaluated by the Teacher Stress Questionnaire would be explored with respect to age and level of teaching. Finally, the job pressure factors, relevant person-facet variables, and moderating factors most correlated to negative stress outcomes, particularly to Free-floating anxiety, Somatic anxiety, Depression, and Overall mental ill-health, would be measured.

After obtaining permission from Professor Cooper for the translation of the Teacher Stress Questionnaire, general guidelines from cross-cultural translation models (Flaherty, Gavia, Pathak, Mitchell, Wintrob, Richman, *et al.*, 1988; Beaton, Bombardier, Guillemin, & Ferraz, 2002) were used as a framework for the translation process. According to these models, the Teacher Stress Questionnaire was translated into Italian by three independent translators. These translators agreed on a common version of the questionnaire, which was then back-translated by an English native speaker who was also an expert in psychology vocabulary. These two versions of the questionnaire were compared and, after further adjustments, an experimental version of the Teacher Stress Questionnaire was developed. Prof. Cooper kindly supplied relevant clarifications concerning the questionnaire's single-item content, which enabled us to obtain a translation as accurate and adequate as possible. This version was initially submitted to a small sample of teachers who differed in age, gender, and level of teaching to verify full comprehension of the single items and the instructions for completion. Once these characteristics were verified and other minor changes were implemented, the current version of the Teacher Stress Questionnaire used in this Italian study was concluded.

METHOD

Participants

The Italian version of the Teacher Stress Questionnaire was individually submitted to 1000 Italian teachers. Multistage sampling was used in the selection of the study sample. Italian teachers were separated into three geographic areas of north, center, and south Italy. From each area schools were selected based on four levels: nursery school, primary school, junior high school, and high school. Teachers were recruited from these schools randomly. To control the quality of data collection, the questionnaire was submitted directly to the teachers during group meetings with psychologists who were familiar with the Teacher Stress Questionnaire. A total of 863 of 1000 questionnaires distributed were returned and considered valid (response rate = 86.3%). Of these, 774 of the teachers (90%) were women; this distribution reflects recent data about the sex distribution of teachers (Fondazione Giovanni Agnelli, 2011). In Italy more than 800,000 of the one million teachers are women. The ages of all teachers considered in the study ranged from 23 to 67 years ($M = 46.5$, $SD = 9.1$).

The teachers were distributed in four levels of teaching and so they were assigned to four groups: nursery school ($n = 146$; 16.9%), primary school ($n = 273$; 31.6%), junior high school ($n = 202$; 23.4%), and high school ($n = 242$; 28%). The teachers were further assigned to three subgroups by age: 23–35 years, 36–50 years, and 51–67 years.

The data were treated to four types of analyses using SPSS-X. Firstly, item analysis was performed measuring the mean, standard deviation, skewness and kurtosis of the items of all single scales of the Teacher Stress Questionnaire. Moreover, inter-item correlations, item-total correlations and Cronbach's α s were checked to evaluate the internal consistency of each scale. Factor analyses were run using the Principal Axis Factoring extraction method with Oblique Promax rotation, to evaluate the construct validity of the Italian version of each scale. Next, the influence of age of teachers and level of teaching on perceived stress dimensions was examined by analysis of variance (ANOVA), considering as dependent variables the subscale scores on each scale. Finally, criterion validity was assessed using multiple regression analysis (stepwise method) considering the subscale scores of the Type A Behavioural Style Inventory, the Job Satisfaction Scale, the Coping Style Inventory, and the Sources of Pressure in Teaching Scale as independent variables and the subscale scores of the Crown-Crisp Experiential Index as dependent variables.

RESULTS

Reliability of the Teacher Stress Questionnaire

With respect to the first aim of the study, this section will present the findings of item analysis (mean, standard deviation, skewness, and kurtosis).

sis) as well as the test for the internal consistency of Teacher Stress Questionnaire single scales (Cronbach's α).

Type A Behavioural Style Inventory.—For this scale, the mean score for the single items varied from a maximum score of 7.54 (Item 1: "Never late") to a minimum score of 5.17 (Item 11: "Hide feelings"), and the mean was 6.43 (SD varied from 1.62 to 2.86). The skewness and kurtosis fell within the range $-1/+1$, indicating normal distributions of the scores; only for Item 1 ("Never late") were the skewness and kurtosis > 1.5 : this item also had the highest mean. These analyses indicate substantially symmetric and normal distributions of the items in the Italian translation of the scale, as required for a multiple regression analysis. Item-total correlations were significant for all items ($r_s > .30$; $p_s < .01$); only two items showed lower r (Item 2: "Very competitive," $r = .26$; and Item 8: "Want good job recognised by others," $r = .28$). In no case was r higher than .80. This finding indicates fundamental item adequacy and consistency with respect to the construct of the whole scale. The mean inter-item correlation was .19, which supports good item discrimination. Cronbach's α was satisfactory at .78. These findings suggest basic adequacy and reliability of the Italian version of the Type A Behavioural Style Inventory.

Job Satisfaction Scale.—For this scale, the mean score for the single items varied from a maximum score of 5.46 (Item 2: "The freedom to choose my method of working") to a minimum score of 3.34 (Item 7: "My rate of pay"), and the mean was 4.68 (SD for single items varied from 1.35 to 1.91). The skewness and kurtosis fell within the range $-1/+1$, indicating normal distributions; for Item 2 ("The freedom to choose my method of working") and Item 6 ("The amount of responsibility you are given"), the skewness and kurtosis were > 1.0 ; these items also had the highest means. Item-total correlations were significant for all the items ($r_s > .30$; $p_s < .01$); only Item 15 ("Your job security") had lower but still acceptable r (.24); in no case was r above .80. The mean inter-item correlation was .34, which indicates a satisfactory item discrimination. Cronbach's α was high at .88. The overall findings indicate excellent adequacy and reliability of the Italian version of the Job Satisfaction Scale.

Sources of Pressure in Teaching Scale.—For this scale, the mean score for the single items varied from a maximum score of 4.44 (Item 43: "The lack of information as to how the changes are to be implemented") to a minimum score of 2.01 (Item 91: "Feeling that apart from teaching I have no other employable skills"), and the mean was 3.31 (SD for the single items varied from 1.29 to 1.93). The skewness and kurtosis were within $-1/+1$, except for Item 63 ("The use of school bells") and for Item 91 ("Feeling that apart from teaching I have no other employable skills"), for which the skewness was < -1.5 : these items had the lowest means. Item-total corre-

lation was significant for all items ($r_s > .30$; $p_s < .01$). Only one item had lower r (Item 3: "Relationship with pupils' parents," $r = .26$); in no case was r higher than .80. The mean inter-item correlation was .26, indicating good item discrimination. Cronbach's α was high at .97. Overall findings support the adequacy and reliability of the Italian version of the Sources of Pressure in Teaching Scale.

Coping Style Inventory.—For this scale, the mean score for the single items varied from a maximum score of 5.09 (Item 4: "Look for ways to make the work more interesting") to a minimum score of 2.30 (Item 24: "Delegation"); the mean was 4.10 (SD for the single items was 0.90–1.60). Skewness and kurtosis were within the range $-1/+1$, indicating normal distributions; for Item 2 ("Try to recognise my own limitations"), Item 4 ("Look for ways to make the work more interesting"), Item 5 ("Reorganise my work") and Item 9 ("Effective time management"), the skewness and kurtosis were > 1.0 : these items had the highest means and standard deviations. Item-total correlation was statistically significant for all items ($r_s > .30$; $p_s < .01$). Seven items had lower r_s : Item 2, "Try to recognise my own limitations," $r = .25$; Item 7, "Resort to hobbies and pastimes," $r = .29$; Item 6, "Seek support and advice from my superiors," $r = .20$; Item 11, "Having a home that is a refuge," $r = .20$; Item 24, "Delegation," $r = .20$; Item 25, "Force one's behaviour and lifestyle to slow down," $r = .20$; and Item 27, "Try to avoid the situation," $r = .20$. In no case was r higher than .80. The mean inter-item correlation was .12, which indicates acceptable item discrimination. Cronbach's α was .78. The analyses indicate basic adequacy and reliability of the Italian version of the Coping Style Inventory.

Construct Validity

This section presents the findings of the factor analyses conducted for each scale and the Cronbach's α calculated for each of the factors extracted. Factor structure was tested using the Principal Axis factor analysis with oblique rotation. The choice of non-orthogonal rotation was based on the assumption that the factors inside each scale could be correlated. Criteria for factor extraction included: eigenvalues greater than 1.0; scree test; parallel analysis of eigenvalues (O'Connor, 2000); item factor loadings ($> .30$), and meaningfulness of the factors. Before the factor structure analysis was conducted, the factorability was assessed by the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity.

Type A Behavioural Style Inventory.—The original inventory consists of 14 items, but in the study reported by Travers and Cooper (1996) there were four factors, including 12 of the original 14 items: Time conscious behaviour (5 items), Ambitious/competitive behaviour (2 items), Efficient behaviour (3 items), and Emotionally suppressive behaviour (2 items).

The present Italian translation used the original questionnaire of 14 items; the assessment of factorability indicated that the KMO measure was .85 and the Bartlett sphericity was significant ($p < .001$), suggesting that the data were adequate for the factor analysis. The parallel analysis and the scree plot suggested that three factors could be extracted. Two items, Item 3 ("Anticipate what others are going to say") and Item 8 ("Want job recognised by others"), were excluded because they had low communalities ($< .35$). Interestingly, these were the same items excluded in the British study. This study's subsequent factor analysis utilised the same 12 items as the British study, but they defined three factors (accounting for 49.33% of the common variance) and not four. In the Italian version, the second factor merged factors 2 and 4 of the British study, and was thus considered an indicator of Emotional suppressive/Ambitious and competitive behaviour (Table 1).

TABLE 1

FACTOR ANALYSIS OF BORTNER'S TYPE A BEHAVIOURAL STYLE INVENTORY: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .77$)

Factor Name and Item	Factor			h^2
	1	2	3	
Factor 1. Time-conscious behaviour				
6. Cerco di fare più cose contemporaneamente [Try to do many things at a time]	.73	-.24	.09	.45
9. Sono veloce (nel mangiare, nel camminare) [Fast (eating, walking)]	.62	-.01	.01	.38
7. Parlo velocemente, con impeto ed enfasi [Emphatic in speech, fast and forceful]	.58	.18	-.20	.40
4. Ho sempre fretta [Always rushed]	.55	.00	-.04	.35
5. Sono impaziente nell'attesa [Impatient while waiting]	.49	.20	-.05	.36
10. Sono esigente [Hard driving]	.34	.28	.07	.35
Factor 2. Emotional suppressive/Ambitious and competitive behaviour				
12. Ho pochi interessi al di fuori di quelli lavorativi e familiari [Few interests outside work/home]	-.11	.59	.09	.35
11. Tendo a nascondere i miei sentimenti [Hide feelings]	.02	.59	-.08	.36
13. Sono ambizioso [Ambitious]	-.05	.59	.08	.36
2. Sono molto competitivo [Very competitive]	.05	.30	-.04	.35
Factor 3. Efficient behaviour				
14. Sono desideroso di realizzare le cose [Eager to get things done]	.14	.06	.68	.61
1. Non sono mai in ritardo [Never late]	-.12	-.01	.34	.35
% Variance explained	29.05	10.90	9.39	
Eigenvalue	3.49	1.31	1.13	

Job Satisfaction Scale.—This consists of 15 items which, in the original study (Travers & Cooper, 1996), referred to five factors: Intrinsic job satisfaction, Extrinsic job satisfaction, Job itself satisfaction, Working conditions satisfaction, and Employee relations satisfaction. In the present study, the assessment of factorability indicated that the KMO measure was .89 and the Bartlett sphericity was significant ($p < .001$), so the data were adequate for the factor analysis. The parallel analysis suggested a two-factor solution, but the scree plot suggested that four factors could be extracted. The four-factor solution was chosen because it was more interpretable and accounted for greater total variance (62.64% of the common variance; Table 2). In the Italian version, the second factor combined Factors 1 (Intrinsic job satisfaction) and 3 (Job itself satisfaction) of the British study, and thus was considered an indicator of Intrinsic and Job itself satisfaction. Notably, a specific aspect that emerged in the Italian school context is that the satisfaction related to employee relations with colleagues and the director constituted the most relevant component of a teacher's perceived job satisfaction.

Sources of Pressure in Teaching Scale.—The original Travers and Cooper inventory consisted of 98 items, but the authors reported in their 1996 study that the factor analysis resulted in 10 factors utilising 55 of the original 98 questions: Pupil/teacher interaction, Management/structure of the school, Class sizes/overcrowding, Changes taking place within education, Appraisal of teachers, Concerns of management, Lack of status/promotion opportunities, "Cover" and staff shortages, Job insecurity, and Ambiguity of the teacher's role. The present study used the 98 items of the original questionnaire; factorability assessment indicated .95 KMO and significant ($p < .001$) Bartlett sphericity, so the data were adequate for factor analysis. The parallel analysis suggested a 10-factor solution, but the scree plot suggested that seven factors could be extracted. Ten-, nine- and eight-factor solutions were difficult to interpret and had several loadings less than .30. The seven-factor solution was chosen because it was more interpretable and accounted for greater total variance. Twenty-four items were found to have low communalities ($< .35$), and were excluded. The results of the seven-factor solution, which utilised 63 items and accounted for 48.48% of the common variance, are reported in Table 3.

When comparing the factors obtained from the Italian sample with those from the original British sample, some similar sources of stress emerged: Pupil/teacher interaction, Cover and staff shortages, Appraisal of teachers/feeling of inadequacy, and Lack of status/professional support. Nevertheless, the present results highlight several sources not prominent in the original study, apparently peculiar to the Italian context: Poor

TABLE 2
FACTOR ANALYSIS OF JOB SATISFACTION SCALE: LOADINGS AND VARIANCE EXPLAINED
(CRONBACH'S $\alpha = .88$)

Factor Name and Item	Factor				h ²
	1	2	3	4	
Factor 1. Employee relations satisfaction					
9. Le relazioni tra la dirigenza e gli insegnanti nella sua scuola [The relations between management and staff in your school]	.80	.18	.15	.18	.73
11. Il modo in cui è diretta la sua scuola [The way your school is managed]	.79	.20	.10	.14	.70
5. Il suo dirigente scolastico [Your director]	.75	.40	.07	.13	.74
3. I suoi colleghi insegnanti [Your fellow teachers]	.38	.37	.10	.17	.35
Factor 2. Intrinsic and job itself satisfaction					
2. La libertà di scegliere il suo personale metodo di lavoro [The freedom to choose your own method of working]	.12	.69	.23	.03	.54
4. Il riconoscimento che riceve per aver fatto un buon lavoro [The recognition you get for good work]	.24	.56	.15	.17	.42
6. Il livello di responsabilità di cui è investito [The amount of responsibility you are given]	.32	.52	.19	.12	.42
8. L'opportunità di mettere a frutto le sue capacità [Your opportunity to use your abilities]	.25	.50	.24	.33	.47
1. Le condizioni fisiche di lavoro [The physical working conditions]	.23	.39	.24	.25	.35
Factor 3. Working conditions satisfaction					
14. Il grado di varietà del suo lavoro [The amount of variety in your job]	.21	.15	.78	.20	.72
13. Il suo orario di lavoro [Your hours of work]	.16	.18	.64	.21	.51
15. La sua sicurezza lavorativa [Your job security]	-.04	.18	.45	-.13	.35
12. L'attenzione con cui le sue proposte vengono considerate [The attention paid to suggestions you make]	.39	.32	.39	.22	.45
Factor 4. Extrinsic job satisfaction					
7. Il livello della sua retribuzione [Your rate of pay]	.10	.17	.04	.70	.53
10. Le sue possibilità di carriera [Your chance of promotion]	.35	.14	.13	.50	.41
% Variance explained	38.14	10.14	7.40	6.97	
Eigenvalue	5.72	1.52	1.11	1.05	

TABLE 3
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (Cronbach's $\alpha = .97$)

Factor Name and Item	Factor							h ²
	1	2	3	4	5	6	7	
Factor 1. Poor working conditions/Lack of resources								
87. Le innumerevoli attività di supervisione che devo svolgere a scuola [The number of supervisory activities I have to perform at school]	.62	-.09	-.16	.03	.18	.08	.02	.41
93. Gli atti vandalici nei locali scolastici [Vandalism of the school premises]	.61	-.16	.42	.03	-.03	-.29	.02	.49
80. La scarsa comunicazione tra i docenti [Poor staff communications]	.59	.00	-.00	.09	-.09	-.03	.04	.37
78. Condizioni di lavoro disagiati [Poor working conditions]	.58	.07	.22	.08	-.10	-.10	-.01	.50
89. L'incertezza circa il livello o la portata delle mie responsabilità [Uncertainty about the degree or area of my responsibility]	.58	-.02	-.11	.06	.09	.10	.07	.44
88. Le aspettative altrui irrealisticamente alte riguardo il mio ruolo [Unrealistically high expectations of others concerning my role]	.55	.03	-.11	-.01	.13	.12	-.01	.39
85. Le richieste inconsuete cui mi trovo di fronte [The unfamiliarity of the demands that I face]	.55	-.12	.07	-.01	-.03	.17	.08	.46 ^a
79. Dovrei gestire una scuola con pochi fondi a disposizione [Having to manage a school on a tight budget]	.52	.22	.02	.01	-.10	-.04	.02	.41
84. La consapevolezza delle condizioni socio-economiche fortemente disagiate degli allievi [Awareness of pupils' social and financial deprivation]	.52	.10	.12	-.14	.05	-.04	.05	.38
94. La mancanza di supporti ausiliari [Lack of auxiliary support]	.51	-.02	.21	-.04	-.05	-.03	.09	.40
83. Portare il lavoro a casa interferisce con la vita familiare [Taking work home interferes with family life]	.50	.05	-.09	.09	.20	-.04	-.01	.35
81. La mancanza di possibilità di carriera [Lack of chance for promotion]	.48	.27	-.11	.23	.05	-.14	-.22	.36

(continued on next page)

(continued on next page)

TABLE 3 (CONT'D)
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
86. L'inadeguata attuazione di cambiamenti nella mia scuola [The inadequate implementation of change in my school]	.46	.04	-.02	.28	-.16	.07	.03	.42
76. La mancanza di tempo per risolvere problemi con i singoli allievi [Lack of time to resolve problems with individual pupils]	.42	.21	-.20	-.20	.02	.23	.09	.46
95. L'integrazione degli allievi con particolari esigenze educative [The integration of pupils with special educational needs]	.41	-.03	.12	-.15	.13	.10	.14	.36
67. Il numero delle interruzioni in classe [The number of interruptions in class]	.39	-.00	.20	-.01	.01	.10	.03	.37
97. Le tensioni razziali all'interno della scuola [Racial tensions within the school]	.37	-.25	.36	.17	-.18	-.07	.08	.38
74. Schemi di lavoro poco definiti [Poorly defined schemes of work]	.30	-.08	.07	.28	-.10	.18	.01	.37
Factor 2. Lack of status / professional support								
29. La mancanza di supporto da parte del Governo [Lack of support from the government]	-.08	.74	.05	-.08	.02	.04	.03	.55
36. La mancanza di supporto da parte del Provveditorato [Lack of support from the school governors]	-.03	.69	.13	-.03	-.10	.10	-.06	.52
28. La mancanza di valore attualmente attribuita all'insegnamento [The lack of value placed on actual teaching itself]	-.13	.58	.23	-.07	.15	.06	.02	.48
20. La scarsa influenza degli insegnanti sulle decisioni che riguardano la scuola nella sua interezza [Teachers can have little influence over school decisions as a whole]	.05	.57	-.03	.20	.10	-.13	.04	.42
38. Il minor rispetto della società nei confronti della mia professione [Society's diminishing respect for my profession]	-.03	.55	.11	-.12	.19	.13	-.04	.45

(continued on next page)

TABLE 3 (CONT'D)
 FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
59. Una retribuzione che non è proporzionata al carico di lavoro [A salary that is out of proportion to workload]	.34	.47	-.10	-.14	.13	-.13	-.15	.35
22. La mancanza di partecipazione ai processi decisionali nella scuola [Lack of participation in decision-making in the school]	.07	.47	-.09	.45	-.02	-.10	.01	.45
26. La mancanza di sostegno da parte del mio sindacato [Lack of support from my union]	-.09	.44	-.02	.31	-.02	.03	.12	.37
43. La mancanza di informazione sul modo in cui attuare le riforme [The lack of information as to how the changes are to be implemented]	.10	.44	-.02	-.15	-.18	.40	.04	.46
33. La mancanza di supporto da parte delle autorità locali [The lack of support from the local authority]	-.01	.42	.28	.12	-.07	-.01	.05	.38
75. L'aumento della pressione esercitata dal Provveditorato [Increasing pressures from school governors]	.31	.36	-.13	-.00	-.09	.34	-.07	.49
Factor 3. Pupil/teacher interaction								
31. Le aggressioni fisiche degli allievi [Physical aggression from pupils]	.01	.06	.87	.03	-.05	-.23	-.00	.60
30. Le aggressioni verbali degli allievi [Verbal aggression from pupils]	-.08	.15	.86	.04	.08	-.17	-.03	.65
34. Assistere all'aumento dell'aggressività tra gli allievi [Witnessing increasing aggression between pupils]	.10	.07	.73	-.10	-.11	-.04	.05	.57
68. Il continuo ribattere degli allievi [The constant "answering back" from pupils]	.06	-.07	.66	.05	-.00	.09	-.07	.55
39. Il numero di conflitti giornalieri che avvengono in classe [The number of daily confrontations in the class]	-.04	-.09	.59	.08	.12	.09	.03	.51
32. La mancanza di sostegno da parte dei genitori su questioni riguardanti la disciplina [The lack of parental back-up on matters of discipline]	.04	.33	.57	-.16	.01	-.05	-.00	.48

(continued on next page)

TABLE 3 (CONT'D)
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (Cronbach's $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
35. Il mancato ricorso a sanzioni nella scuola [No recourse to sanctions in the school]	-.09	.11	.50	.14	-.05	.13	-.04	.38
53. Insegnare a persone che non attribuiscono valore all'istruzione [Teaching those who do not value education]	.16	.12	.44	-.13	.09	.16	-.10	.46
54. Insegnare a chi dà le cose per scontate [Teaching those who take things for granted]	.14	.08	.37	.03	.06	.20	-.15	.41
57. Mantenere la disciplina [Maintaining discipline]	-.09	.01	.33	.06	.30	.22	-.02	.45
58. Gli allievi che cercano continuamente di "metterli alla prova" [When pupils try to "test" you all the time]	-.02	-.08	.30	.21	.25	.22	-.07	.47
Factor 4. Management / Lack of decision latitude								
91. L'atteggiamento dei genitori circa la mia adesione alle politiche del sindacato (ad es. scioperi) [Parental attitudes towards my adherence to union policies, e.g., strikes]	.05	-.29	-.00	.62	-.01	.07	.02	.44
23. La natura "gerarchica" della mia struttura scolastica [The "hierarchical" nature of the structure of my school]	-.11	.33	-.12	.59	.07	-.06	.07	.45
63. L'uso delle campane scolastiche [The use of school bells]	-.05	-.25	-.05	.58	.08	.15	-.02	.41
62. La mia scuola è troppo "tradizionalista" e non si muove al passo con i tempi [My school is too "traditional" and is slow to move with the times]	.07	-.02	.10	.53	-.15	.06	-.01	.37
82. Avere troppo poche responsabilità all'interno della scuola [Too little responsibility within the school]	.31	.02	.10	.50	.03	-.11	-.16	.37
27. La mancanza di chiarezza circa il mio ruolo all'interno della scuola [The lack of clarity concerning my role within the school]	-.14	.17	.15	.49	.00	-.02	.19	.44

(continued on next page)

TABLE 3 (CONT'D)
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (CROBACH'S $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
21. I conflitti tra il mio e gli altri Istituti per l'acquisizione delle risorse [Conflict between my department and others for resources]	.11	.25	-.10	.48	.06	-.06	.08	.41
61. Regiare in modo troppo personale alle critiche degli allievi [Reacting too personally to pupils' criticism]	-.12	-.17	.25	.44	.06	.28	-.01	.51
25. Il conflitto tra i bisogni della mia classe e le opinioni dello staff dirigenziale [Conflict between the needs of my department/class and the views of senior management]	.08	.19	.09	.40	-.04	-.08	.17	.38
60. Non avere abbastanza opportunità di decidere in prima persona [Not enough opportunity to make my own decisions]	.21	.20	-.00	.39	-.01	.16	-.17	.43
49. La mancanza di supporto da parte del Dirigente Scolastico. [Lack of support from the Head teacher]	.06	.11	.12	.33	-.20	.22	.03	.35
Factor 5. Emotional involvement with pupils and parents								
4. Il coinvolgimento emotivo troppo alto nel rapporto con gli allievi [Over-emotional involvement with pupils]	.03	.03	.00	-.09	.71	-.09	.10	.46
5. Gestire problemi comportamentali [Dealing with behavioural problems]	.14	-.03	.08	-.26	.67	-.05	.10	.47
1. Costruire e mantenere le relazioni con gli allievi [Building and maintaining relationships with pupils]	-.01	.10	-.03	-.03	.66	-.07	-.02	.38
3. Le relazioni con i genitori dei miei allievi [Relationships with pupils' parents]	-.02	.04	-.08	-.04	.63	.00	.03	.36
15. La mancanza di momenti di "assenza di contatto" [Lack of "non contact" time]	-.01	-.14	-.01	.22	.55	.06	-.00	.41
19. Dover partecipare agli incontri con i genitori [Having to attend parents' evenings]	-.07	.14	-.05	.21	.41	.12	.05	.38

(continued on next page)

TABLE 3 (CONT'D)
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
11. Dover prendere continue decisioni in classe [The need for constant decision-making in the classroom]	-.03	.05	.12	.24	.41	-.05	.08	.38
56. Dover continuamente instaurare nuove relazioni [Continually having to form new relationships]	.20	-.05	.08	.16	.37	.10	-.08	.42
55. Doversi occupare di allievi che richiedono un'attenzione immediata [Dealing with children who demand immediate attention]	.11	.01	.15	.07	.31	.24	-.06	.43
Factor 6. Appraisal of teachers/Feelings of inadequacy								
46. L'elevata importanza attribuita nella scuola ai risultati conseguiti [Academic pressure within the school]	.03	.06	-.14	.04	.05	.67	-.07	.44
48. Le elevate aspettative di buoni risultati dei genitori [High demands from parents for good results]	.01	.17	-.02	-.06	-.02	.62	.04	.45
42. Il fatto che si vada verso una programmazione nazionale [The move towards a "National Curriculum"]	-.04	.21	-.05	.09	-.06	.57	-.01	.40
51. I casi in cui la mia prestazione è valutata da altri [When my performance is assessed by others]	-.01	-.06	-.08	.23	.01	.55	.00	.39
41. I continui cambiamenti che avvengono nella mia professione [The constant changes taking place within the profession]	.02	.30	-.01	-.09	-.09	.44	.12	.38
71. Insegnare tenendo conto degli standard di esame [Teaching to exam standard]	.04	-.05	.04	.12	.10	.43	-.00	.35
72. L'avanzamento di carriera ha condotto ad una riduzione dei rapporti con gli allievi in classe [Promotion has led to few class contacts with pupils]	.27	-.10	-.03	.09	.02	.40	-.01	.36
50. Avere la sensazione che la mia preparazione non sia adeguata [Feeling that my training is not appropriate]	.02	-.13	.03	.31	-.06	.40	.07	.35

(continued on next page)

TABLE 3 (CONT'D)
FACTOR ANALYSIS OF SOURCES OF PRESSURE IN TEACHING SCALE: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .97$)

Factor Name and Item	Factor							h^2
	1	2	3	4	5	6	7	
44. Dover essere un "tuttofare, maestro di nessuno" [Having to be a "jack of all trades, master of none"]	.24	.32	-.04	-.05	-.11	.35	.06	.46
47. L'aumento del coinvolgimento in problemi "pastorali" [Increasing involvement with "pastoral" issues]	.13	.09	.05	.04	.19	.34	-.04	.39
Factor 7. "Cover" and staff shortages								
7. Non sapere quando dovrò sostituire qualcuno [The unpredictability of "cover" periods]	.03	-.02	-.06	.03	.02	-.01	.79	.61
8. I casi in cui la sostituzione di colleghi assenti rende la classe troppo numerosa [When "cover" for absent colleagues leads to "large" classes]	.08	.00	.01	-.01	.06	-.01	.72	.58
6. Dover sostituire colleghi che insegnano discipline a me non familiari [Having to "cover" in unfamiliar areas of the curriculum]	.06	-.03	-.05	.06	.09	.00	.67	.51
9. L'impossibilità di pianificare in anticipo a causa dei continui mutamenti [Inability to plan ahead due to constant changes]	.09	.04	-.01	.07	.06	.00	.55	.42
% Variance explained	26.75	5.49	4.11	3.97	3.27	2.67	2.22	
Eigenvalue	19.79	4.06	3.05	2.94	2.42	1.97	1.64	

working conditions / Lack of resources, Emotional involvement with pupils and parents, and Management / Lack of decision latitude.

Coping Style Inventory.—This scale consists of 28 items from the eight factors in the original study: Prioritise / objective coping, Hobbies and pastimes, Mobilised social support, Time measures, Innovation, Suppression of stress, Non-confronting of the situation, and Non-involvement and delegation. In the present study, the factorability assessment yielded a .81 KMO measure and significant ($p < .001$) Bartlett sphericity, so the data were adequate for the factor analysis. The parallel analysis suggested a five-factor solution, while scree plot suggested eight. The latter solution was chosen because it was more interpretable and congruous with the original scale's structure, and because it accounted for greater total variance. The results of the eight-factor solution, accounting for 54.48% of the common variance, are reported in Table 4.

When comparing the coping factors obtained from the Italian sample with those obtained from the British sample, several coping factors were confirmed in the Italian setting: Innovation, Suppression of stress, Hobbies and pastimes, and Mobilised social support. In the British study, Factor 1 (Prioritize/objective coping) referred to problem-focused coping modalities connected to selective attention, setting of priorities, and objective coping and Factor 4 referred to Time measures, whereas in the present study Factor 1 (Coping focused on the problem and time management) combined these coping modalities. Moreover, in the present study both Selection of priorities and Objective coping constituted independent modalities of coping (Factor 5 and Factor 8), and the non-confronting modalities of coping emerged as related to tendencies to delegation (factor 3: Non-confronting of the situation / Delegation).

Differences in Teacher Stress Questionnaire by Age and Teaching Level

According to Travers and Cooper's approach, stress is considered a dynamic relational concept which depends on the constant interplay between individual and situational factors. In this study, the differences in the Teacher Stress Questionnaire dimensions were explored with respect to two demographic characteristics that are considered to influence job attitudes and work-related stress dimensions: the personal characteristic of the teacher's age (Smith, *et al.*, 2000; Ng & Feldman, 2010; Shultz, Wang, Crimmins, & Fisher, 2010; Mauno, Ruokolainen, & Kinnunen, 2013), and a situational characteristic, the level of teaching (Travers & Cooper, 1996). This section presents, for each scale of the Teacher Stress Questionnaire, the findings from ANOVAs by age and level of teaching (dependent variables: factor scores, total and subscale scores of each scale; *post hoc* comparisons with Bonferroni's correction).

Teachers between 23 and 35 years old reported significantly higher scores than the other subgroups of teachers with respect to the Job satis-

TABLE 4
FACTOR ANALYSIS OF COPING STYLES INVENTORY: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .78$)

Factor Name and Item	Factor								h ²
	1	2	3	4	5	6	7	8	
Factor 1. Coping focused on the problem and time management									
21. Stabilisco delle priorità in base alle quali gestisco i problemi [Set priorities and deal with problems accordingly]	.70	-.21	.01	-.11	.08	.03	.00	.12	.47
15. Pianifico in anticipo [Plan ahead]	.62	.19	.04	.03	-.05	-.03	.01	-.13	.46
22. Cerco di guardare la situazione dall'esterno, esaminandola attentamente [Try to "stand aside" and think through the situation]	.57	-.12	.00	.03	.13	-.02	.04	-.02	.36
9. Cerco di gestire efficacemente il tempo [Effective time management]	.49	.26	-.02	-.06	-.13	.07	-.06	.19	.47
1. Gestisco i problemi immediatamente man mano che si presentano [Deal with the problems immediately as they occur]	.37	.14	.07	.05	-.01	-.09	-.05	.16	.35
26. Accetto la situazione e trovo il modo di convivervi [Accept the situation and learn to live with it]	.30	-.11	.12	.12	.04	-.03	.08	.23	.35
Factor 2. Innovation									
5. Riorganizzo il mio lavoro [Reorganise my work]	-.10	.67	.05	.05	.11	-.02	.07	.04	.43
4. Cerco di trovare un modo per rendere il lavoro più interessante [Look for ways to make the work more interesting]	-.06	.60	.02	-.03	.08	.12	.03	.13	.44
Factor 3. Non confrontive of the situation and delegation									
24. Delego [Delegation]	.04	-.15	.64	.09	-.09	-.05	.01	-.04	.50
25. Mi impongo un rallentamento del comportamento e dello stile di vita [Force one's behaviour and lifestyle to slow down]	.16	-.05	.47	-.09	-.03	.12	.00	.04	.35
27. Cerco di evitare la situazione [Try to avoid the situation]	-.02	-.08	.45	.16	.05	-.09	.11	-.02	.35
(continued on next page)									

(continued on next page)

TABLE 4 (CONT'D)
FACTOR ANALYSIS OF COPING STYLES INVENTORY: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .78$)

Factor Name and Item	Factor								h^2
	1	2	3	4	5	6	7	8	
6. Cerco sostegno e consiglio presso i miei superiori [Seek support and advice from my superiors]	.04	.21	.39	.07	-.06	-.02	.10	-.12	.35
28. Cerco il maggior sostegno sociale possibile [Seek as much social support as possible]	-.08	.20	.36	.06	.06	.10	-.02	-.03	.35
Factor 4. Hobbies and pastimes									
17. Espando i miei interessi e le miei attività extra-lavorative [Expand interest and activities outside work]	.04	.05	.09	.68	.06	-.13	-.14	-.03	.51
7. Mi rifugio in hobby e passatempo [Resort to hobbies and pastimes]	-.14	-.05	.11	.66	-.10	.06	.01	.13	.44
14. Cerco di tenermi occupato ["Stay busy"]	.25	.10	-.13	.36	-.06	.02	.17	-.12	.35
Factor 5. Selection and priorities									
16. Cerco di non farmi sommeregere dai problemi e di allentare la tensione [Not "bottling things up" and being able to release energy]	.19	.08	.02	-.09	.69	-.14	-.12	-.05	.61
19. Uso l'attenzione selettiva (mi concentro su specifici problemi) [Use selective attention (concentrating on specific problems)]	.05	.10	-.03	.03	.43	.09	.03	.04	.35
13. Mantengo deliberatamente separati il lavoro e la vita privata [Deliberately separate "home" and "work"]	.11	.07	-.08	.03	.41	-.09	.08	-.05	.35
3. Provo a guadagnare tempo e a dilazionare il problema ["Buy time" and stall the issue]	-.18	.11	.20	-.05	.30	.09	.22	.10	.35
Factor 6. Mobilised social support									
12. Mi confido con amici che possano capirmi [Talk to understanding friends]	.01	.11	.19	-.15	-.03	.62	-.10	-.06	.43
11. Trovo rifugio in casa mia [Having a home that is a refuge]	-.06	-.02	-.13	.11	-.09	.49	.18	.00	.35

(continued on next page)

TABLE 4 (CONT'D)
FACTOR ANALYSIS OF COPING STYLES INVENTORY: LOADINGS AND VARIANCE EXPLAINED (CRONBACH'S $\alpha = .78$)

Factor Name and Item	Factor								<i>h</i> ²
	1	2	3	4	5	6	7	8	
20. Cerco di distrarmi (per tenere la mente libera dai problemi) [Use distraction (to take your mind off things)]	.15	-.15	-.05	.21	.18	.23	.14	.03	.30
Factor 7. Suppression of stress									
10. Reprimo le emozioni e mi sforzo di nascondere lo stress [Suppress emotions and try not to let the stress show]	.00	.08	.09	-.04	-.02	.04	.56	.06	.35
23. Ricorro a regole e procedure di comportamento [Resort to rules and regulations]	.42	-.01	.06	-.10	.06	.06	.43	-.12	.39
18. Mi giovo di stabili relazioni interpersonali [Have stable relationships]	.18	.01	-.06	.32	.14	.19	.33	-.03	.39
Factor 8. Objective coping									
2. Cerco di riconoscere i miei limiti [Try to recognise my own limitations]	.03	.24	-.07	.03	-.04	.02	.01	.53	.40
8. Cerco di affrontare la situazione oggettivamente in modo non emotivo [Try to deal with the situation objectively in an unemotional way]	.18	.30	-.06	.03	.01	-.15	.08	.46	.48
% Variance explained	17.55	9.83	5.62	5.09	4.54	4.12	3.91	3.82	
Eigenvalue	4.91	2.75	1.57	1.43	1.27	1.15	1.09	1.07	

faction scale's Factor 1, Employee relations satisfaction ($F_{2,792} = 3.15, p < .05$) and Factor 4, Extrinsic job satisfaction ($F_{2,792} = 4.55, p < .05$), and the Coping Style Inventory's Factor 3, Non-confronting of the situation and delegation ($F_{2,611} = 3.16, p < .05$); they reported lower scores than the other two age groups of teachers with respect to the Job satisfaction scale's Factor 3, Working conditions satisfaction ($F_{2,795} = 11.18, p < .001$), and with respect to the Coping Style Inventory's Factor 8, Objective coping ($F_{2,785} = 9.94, p < .001$); they also reported significantly lower scores than teachers between 51 and 67 years old with respect to the Coping Style Inventory's Factor 1, Coping focused on the problem and time management ($F_{2,761} = 4.61, p < .05$), significantly lower scores than teachers between 36 and 50 years old with respect to the Source of pressure in the teaching scale's Factor 2, Lack of status / professional support ($F_{2,662} = 3.82, p < .05$), and significantly higher scores than teachers between 51 and 67 years old with respect to Factor 4, Management / Lack of decision latitude ($F_{2,649} = 3.58, p < .05$). Teachers between 36 and 50 years old reported significantly higher scores than teachers between 23–35 years old with respect to the Source of pressure in the teaching scale's Factor 2, Lack of status / Professional support ($F_{2,662} = 3.82, p < .05$). Teachers between 51 and 67 years old reported significantly higher scores than the other two subgroups with respect to the Job satisfaction scale's Factor 3, Working conditions satisfaction ($F_{2,795} = 11.18, p < .001$), and lower scores than the other two subgroups with respect to the Source of pressure in the teaching scale's Factor 4, Management / Lack of decision latitude ($F_{2,649} = 3.58, p < .05$); they reported significantly higher scores than teachers between 23 and 35 years old with respect to the Coping Style Inventory's Factor 1, Coping focused on the problem and time management ($F_{2,761} = 4.61, p < .05$), and also Factor 8, Objective coping ($F_{2,785} = 9.94, p < .001$); and they also reported significantly lower scores than teachers between 23 and 35 years old with respect to the Job satisfaction scale's Factor 4, Extrinsic job satisfaction ($F_{2,798} = 4.55, p < .05$) and with respect to the Coping Style Inventory's Factor 3, Non-confrontive of the situation / Delegation ($F_{2,611} = 3.16, p < .05$).

When comparing subgroups of teachers by level of teaching, it was found that nursery school teachers reported significantly lower scores than junior high school teachers with respect to the Sources of pressure in the teaching scale's Factor 5, Emotional involvement with pupils and parents ($F_{3,660} = 4.02, p < .01$), to the Coping Style Inventory's Factor 3, Non-confrontive of the situation / Delegation ($F_{3,621} = 13.58, p < .001$), and to the Coping Style Inventory's Factor 4, Hobbies and pastimes ($F_{3,790} = 7.95, p < .001$); moreover, they reported significantly higher scores than junior high school teachers with respect to the Coping Style Inventory's Factor 2, Innovation ($F_{3,809} = 23.93, p < .001$), and Factor 8, Objective coping

($F_{3,807} = 13.06, p < .001$). Finally, they reported significantly lower scores than primary school teachers on the Crown-Crisp Experiential Index scores of Free-floating anxiety ($F_{3,777} = 6.55, p < .001$).

Primary school teachers reported significantly higher scores than junior high and high school teachers with respect to the Coping Style Inventory's Factor 2, Innovation ($F_{3,809} = 23.93, p < .001$), and with respect to the Crown-Crisp Experiential Index scores of Somatic concomitants of anxiety ($F_{3,722} = 11.24, p < .001$); moreover, they reported significantly lower scores than junior high school teachers with respect to the Coping Style Inventory's Factor 3, Non-confrontive of the situation/Delegation ($F_{3,621} = 13.58, p < .001$), Factor 4, Hobbies and pastimes ($F_{3,790} = 7.95, p < .001$), and Factor 8, Objective coping ($F_{3,807} = 13.06, p < .001$); and with respect to the Job Satisfaction Scale's Factor 1, Employee relations satisfaction ($F_{3,815} = 2.90, p < .05$). Moreover, they reported significantly higher scores than high school teachers with respect to Sources of pressure in the teaching scale's Factor 1, Poor working conditions/Lack of resources ($F_{3,657} = 6.43, p < .001$), Factor 6, Appraisal of teachers/Feelings of inadequacy ($F_{3,668} = 4.32, p < .01$), and Factor 7, "Cover" and staff shortages ($F_{3,717} = 11.81, p < .001$), and with respect to the Crown-Crisp Experiential Index scores on Free-floating anxiety ($F_{3,777} = 6.55, p < .001$), Depression ($F_{3,756} = 5.48, p < .02$), and Overall mental ill-health ($F_{3,638} = 7.21, p < .001$).

Junior high school teachers reported significantly higher scores than high school teachers with respect to the Sources of pressure in the teaching scale's Factor 7, "Cover" and staff shortages ($F_{3,717} = 11.81, p < .001$), the Crown-Crisp Experiential Index scores on Somatic anxiety ($F_{3,722} = 11.24, p < .001$), and the Coping Style Inventory's Factor 3, Non-confrontive of situation/Delegation ($F_{3,621} = 13.58, p < .001$), and lower scores than high school teachers on the Coping Style Inventory's Factor 2, Innovation ($F_{3,809} = 23.93, p < .001$), and Factor 8, Objective coping ($F_{3,807} = 13.06, p < .001$).

Teacher Stress Questionnaire Variables Associated With Psychological Strain

The fourth goal of the present study presents the findings of regression analyses (stepwise method) of the independent variables (Type A behaviour style, job stressors, job satisfaction, coping style, demographics) against the dependent variables (Overall mental ill-health, Free-floating anxiety, Somatic concomitants of anxiety, and Depression measured by the Crown-Crisp Experiential Index). The regression models showed (Table 5) that a dispositional factor, the tendency toward Time-conscious behaviour (Factor 1 of the Type A Behavioural Style Inventory), predicted higher scores on all dimensions of psychological strain: Overall mental ill-health, Free-floating anxiety, Somatic anxiety, and Depression; and that a situational factor, the Poor working conditions/Lack of resources (Fac-

TABLE 5
MULTIPLE REGRESSION ANALYSIS ON TOTAL MENTAL HEALTH, FREE-FLOATING ANXIETY, SOMATIC ANXIETY, AND DEPRESSION

Step	Variable	<i>F</i>	Multiple <i>R</i> ²	β	<i>t</i>
Multiple regression analysis on Total Mental Health					
1	Job Satisfaction Factor 2				
	Intrinsic and job itself satisfaction	41.43‡	.14	-.29	-5.1‡
2	Type A Behaviour Factor 1				
	Time-conscious behaviour	44.81‡	.27	.29	5.5‡
3	Source Factor 1				
	Poor working conditions and lack of resources	36.02‡	.31	.15	2.9*
4	Coping Factor 1				
	Coping focused on the problem and time management	30.23‡	.33	-.25	-4.0‡
5	Coping Factor 7				
	Suppression of stress	26.36‡	.35	.19	3.1*
6	Job Satisfaction Factor 4				
	Extrinsic job satisfaction	23.38‡	.37	-.14	-2.4*
7	Coping Factor 4				
	Hobbies and pastimes	20.91‡	.38	-.11	-2.0*
Multiple regression analysis on Free-floating Anxiety					
1	Type A Behaviour Factor 1				
	Time-conscious behaviour	42.14‡	.13	.31	6.0‡
2	Job Satisfaction Factor 2				
	Intrinsic and job itself satisfaction	42.43‡	.23	-.26	-4.6‡
3	Coping Factor 1				
	Coping focused on the problem and time management	33.59‡	.27	-.31	-5.1‡
4	Coping Factor 7				
	Suppression of stress	28.80‡	.29	.20	3.4*
5	Job Satisfaction Factor 4				
	Extrinsic job satisfaction	24.14‡	.30	-.12	-2.0*
Multiple regression analysis on Somatic Anxiety					
1	Source Factor 1				
	Poor working conditions and lack of resources	31.63‡	.11	.22	4.1‡
2	Job Satisfaction Factor 2				
	Intrinsic and job itself satisfaction	29.91‡	.19	-.28	-5.3‡
(continued on next page)					

* $p < .05$. ‡ $p < .001$.

TABLE 5 (CONT'D)
MULTIPLE REGRESSION ANALYSIS ON TOTAL MENTAL HEALTH, FREE-FLOATING ANXIETY, SOMATIC ANXIETY, AND DEPRESSION

Step	Variable	<i>F</i>	Multiple <i>R</i> ²	β	<i>t</i>
3	Type A Factor 1				
	Time-conscious behaviour	26.33‡	.23	.22	4.1‡
4	Coping Factor 1				
	Coping focused on the problem and time management	22.04‡	.25	-.31	-4.3‡
5	Coping Factor 5				
	Selection and priorities	19.45‡	.27	.14	2.2*
6	Coping Factor 8				
	Objective coping	17.21‡	.29	.12	1.9*
7	Coping Factor 7				
	Suppression of stress	15.57‡	.30	.15	2.4*
8	Coping Factor 4				
	Hobbies and pastimes	14.31‡	.31	-.11	-2.0*
Multiple regression analysis on Depression					
1	Job Satisfaction Factor 2				
	Intrinsic and job itself satisfaction	39.98‡	.13	-.26	-4.5‡
2	Type A Behaviour Factor 1				
	Time-conscious behaviour	40.53‡	.23	.28	5.3‡
3	Source Factor 1				
	Poor working conditions and lack of resources	32.49‡	.26	.16	3.0*
4	Coping Factor 6				
	Mobilised social support	25.70‡	.28	.14	2.6*
5	Coping Factor 1				
	Coping focused on the problem and time management	22.36‡	.29	-.16	-2.9*
6	Job Satisfaction Factor 4				
	Extrinsic job satisfaction	19.76‡	.31	-.13	-2.2*

**p* < .05. ‡*p* < .001.

tor 1 of the Sources of pressure in teaching scale), predicted significantly higher Overall mental ill-health, Somatic anxiety, and Depression scores. Moreover, the analyses emphasized the importance of several coping strategies and job satisfaction factors which were able to predict the severity and quality of perceived psychological strain, and thus confirmed the moderating role exerted by these aspects within the stress process as conceived and measured by the Teacher Stress Questionnaire.

DISCUSSION

The findings of this study conducted through the Italian version of the Teacher Stress Questionnaire revealed adequate validity and reliability of this measure for the analysis of teacher stress in Italian school settings. From the factor analysis of each of the scales of the Teacher Stress Questionnaire, several consistent and psychologically meaningful factors emerged related to the sources of pressure in the school setting, the coping strategies adopted, the job satisfaction perceived, and the severity and quality of teachers' psychological strain. Moreover, the study has highlighted some dimensions of stress that are specific to the Italian school setting, also with respect to age and level of teaching. However, the multiple regression analyses pointed out some important variables tested by the questionnaire, e.g., dispositional factors such as Type A behaviours and moderating factors such as the coping strategies used and the job satisfaction perceived. From this perspective, the multiple regression analyses highlighted the balance among dispositional, situational, and moderating factors within Travers and Cooper's conceptualization of stress upon which the Teacher Stress Questionnaire is based. Finally, a reliability analysis of the questionnaire, tested by means of the internal consistency evaluation of the single scales, also revealed basic adequacy.

To extend these promising results, the individual and situational factors identified in the context of the transactional model of stress should be more deeply explored in Italian school settings. Further research may use the Italian version of the Teacher Stress Questionnaire to examine the interaction of stress dimensions measured by the questionnaire with respect to demographic characteristics such as teacher's age and level of teaching, and to explore the influence of each dimension of the model on physical and psychological health conditions perceived by teachers. Despite its limitations, this report confirmed that a psychometrically valid version of the Teacher Stress Questionnaire is available and can be used efficiently as evidence-based input to health-promoting measures of organisational and personnel development in Italian school settings.

REFERENCES

- BAUER, J., STAMM, A., VIRNICH, K., WISSING, K., KRISTON, L., MÜLLER, U., WENG, G., STÜRM-LINGER, R., WIRSCHING, M., & SCHAARSCHMIDT, U. (2006) Correlation between burn-out syndrome and psychological and psychosomatic symptoms among teachers. *International Archives of Occupational Health*, 79, 199-204.
- BEATON, D. E., BOMBARDIER, C., GUILLEMIN, F., & FERRAZ, M. B. (2002) *Recommendations for the cross-cultural adaptation of health status measures*. Rosemont, IL: American Academy of Orthopedic Surgeons.
- BENMANSOUR, N. (1998) Job satisfaction, stress and coping strategies among Moroccan high school teachers. *Mediterranean Journal of Educational Studies*, 3, 13-33.

- BORTNER, R. W. (1969) A short rating scale as a potential measure of Pattern A behaviour. *Journal of Chronic Diseases*, 22, 87-91.
- BYRNE, B. M. (1999) Teacher burnout: the nomological network. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout*. New York: Cambridge Univer. Press. Pp. 15-37.
- CHAN, D. W., & HUI, E. K. P. (1995) Burnout and coping among Chinese secondary school teachers in Hong Kong. *British Journal of Educational Psychology*, 65, 15-25.
- COOPER, C. L., SLOAN, S. J., & WILLIAMS, S. (1988) *Occupational stress indicator*. Windsor, UK: NFER Nelson.
- CROWN, S., & CRISP, A. H. (1979) *Manual of the Crown-Crisp Experiential Index*. London, UK: Hodder & Stoughton.
- DEWE, P., LEITER, M., & COX, T. (Eds.) (2000) *Coping, health and organizations*. London, UK: Taylor & Francis.
- DUNHAM, J., & VARMA, V. (Eds.) (1998) *Stress in teachers: past, present and future*. London, UK: John Wiley and Sons.
- FARBER, B. A. (1991) *Crisis in education: stress and burnout in the American teacher*. San Francisco, CA: Jossey-Bass.
- FLAHERTY, J. A., GAVIRIA, F. M., PATHAK, D., MITCHELL, T., WINTROB, R., RICHMAN, J. A., & BIRZ, S. (1988) Developing instruments for cross-cultural psychiatric research. *Journal of Nervous & Mental Disease*, 176(5), 257-263.
- FONDAZIONE GIOVANNI AGNELLI. (2011) *Rapporto sulla scuola in Italia 2011*. Roma-Bari, Italy: Laterza.
- FRIEDMAN, M., & ROSENMAN, R. H. (1959) Association of specific overt behavior pattern with increases in blood cholesterol, blood clotting time, incidence of *arcus senilis* and clinical coronary artery disease. *Journal of the American Medical Association*, 169, 1286-1296.
- GUGLIELMI, R. S., & TATROW, K. (1998) Occupational stress, burnout and health in teachers: a methodological and theoretical analysis. *Review of Educational Research*, 68, 61-99.
- KYRIACOU, C. (2001) Teacher stress: directions for future research. *Educational Review*, 53, 27-35.
- LAZARUS, R. S. (1999) *Stress and emotion: a new synthesis*. London, UK: Free Association Books.
- MAUNO, S., RUOKOLAINEN, M., & KINNUNEN, U. (2013) Does aging make employees more resilient to job stress? Age as a moderator in the job stressor-well-being relationship in three Finnish occupational samples. *Aging & Mental Health*, 17(4), 411-422.
- NG, T. W. H., & FELDMAN, D. C. (2010) The relationships of age with job attitudes: a meta-analysis. *Personnel Psychology*, 63, 677-718.
- O'CONNOR, B. P. (2000) SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396-402.
- PITHERS, R. T., & SODEN, R. (1998) Scottish and Australian teacher stress and strain: a comparative study. *British Journal of Educational Psychology*, 68, 269-279.
- SHULTZ, S. K., WANG, M., CRIMMINS, E. M., & FISHER, G. G. (2010) Age differences in the Demand-Control Model of work stress: an examination of data from 15 European Countries. *Journal of Applied Gerontology*, 29(1), 21-47.

- SMITH, A., BRICE, C., COLLINS, A., MATTHEWS, V., & MCNAMARA, R. (2000) *The Scale of Occupational Stress: a further analysis of the impact of demographic factors and type of job*. Cardiff, UK: Health and Safety Executive.
- SUTHERLAND, V. J., & COOPER, C. L. (1988) Sources of work stress. In J. J. Hurrell, Jr., L. R. Murphy, S. L. Sauter, & C. L. Cooper (Eds.), *Occupational stress: issues and developments in research*. London, UK: Taylor & Francis.
- TRAVERS, C. J., & COOPER, C. L. (1996) *Teachers under pressure: stress in the teaching profession*. London, UK: Routledge.
- UNTERBRINK, T., HACK, A., PFEIFER, R., BUHL-GRIEBHABER, V., MÜLLER, U., WESCHE, H., FROMMHOLD, M., SCHEUCH, K., SEIBT, R., WIRSCHING, M., & BAUER, J. (2007) Burnout and effort-reward-imbalance in a sample of 949 German teachers. *International Archives of Occupational and Environmental Health*, 80, 433-441.
- VAN DICK, R., & WAGNER, U. (2001) Stress and strain in teaching: a structural equation approach. *British Journal of Educational Psychology*, 71, 243-259.
- WARR, P., COOK, J., & WALL, T. (1979) Scales for the measurement of some work attitudes and aspects of psychological well-being. *The Journal of Occupational Psychology*, 52, 129-148.
- ZURLO, M. C. (2013) Dimensioni dello stress degli insegnanti Italiani e Tedeschi: studio comparativo con il modello Effort-Reward Imbalance [Stress dimensions of Italian and German teachers: a comparative study with the Effort-Reward Imbalance Model]. In E. Schafroth, M. Nicklaus, C. Schwarzer, & D. Conte (Eds.), *Italien, Deutschland, Europa: kulturelle identitäten und interdependenzen* [Italy, Germany, Europe: cultural identities and interdependencies]. Oberhausen, Germany: Athena. Pp. 233-251.
- ZURLO, M. C., PES, D., & COOPER, C. L. (2007) Stress in teaching: a study of occupational stress and its determinants among Italian schoolteachers. *Stress & Health*, 23(4), 231-241.
- ZURLO, M. C., PES, D., & SIEGRIST, J. (2010) Validity and reliability of the Effort-Reward Imbalance Questionnaire in a sample of 673 Italian teachers. *International Archives of Occupational and Environmental Health*, 83, 665-674.
- ZURLO, M. C., & PES, D. (2012) Effort-Reward Imbalance e salute psicofisica degli insegnanti. *Giornale Italiano di Medicina del Lavoro ed Ergonomia (Supplemento B, Psicologia)*, 34(2), 81-87.

Accepted September 25, 2013.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Mental &
Physical
Health*



NATIONAL INTELLIGENCE ESTIMATES AND THE FAILED STATE INDEX¹

MARTIN VORACEK

University of Vienna
University of Göttingen

Summary.—Across 177 countries around the world, the Failed State Index, a measure of state vulnerability, was reliably negatively associated with the estimates of national intelligence. Psychometric analysis of the Failed State Index, compounded of 12 social, economic, and political indicators, suggested factorial unidimensionality of this index. The observed correspondence of higher national intelligence figures to lower state vulnerability might arise through these two macro-level variables possibly being proxies of even more pervasive historical and societal background variables that affect both.

The publication of a compilation of national intelligence (IQ) estimates for the world's countries by Lynn and Vanhanen (2002; revised and extended versions: Lynn & Vanhanen, 2006, 2012a) has spawned considerable interest among researchers across a variety of scientific disciplines. A review of the emerging interdisciplinary literature using these national IQ figures has been published recently (Lynn & Vanhanen, 2012b).

A cited reference search (as of April 2013) for the first two (of three) source books in Google Scholar indicated that the first book (Lynn & Vanhanen, 2002) has been cited almost 450 times and the second book (Lynn & Vanhanen, 2006) more than 200 times. Content analysis of the results of this cited reference search suggests that the research interest in these national IQ data is primarily found in psychology (foremost, intelligence research), economics, and anthropology, with a number of further accounts coming from public health science, demography, and other social science disciplines, such as biomedical social science.

Although the sources and procedures used by Lynn and Vanhanen (2002, 2006) for compiling the national IQ estimates have been criticized from various points of view (e.g., Barnett & Williams, 2004), there also has been evidence emerging for the reliability, validity and, thus, applicability of these data (e.g., Dickerson, 2006; Lynn, Meisenberg, Mikk, & Williams, 2007; Rindermann, 2007). Up to now, across dozens of studies,

¹Address correspondence to Martin Voracek, D.Sc., D.M.Sc., Ph.D., Department of Basic Psychological Research and Research Methods, School of Psychology, University of Vienna, Liebiggasse 5, Room 03-46, A-1010 Vienna, Austria, or Georg Elias Müller Department of Psychology, Georg August University of Göttingen, Goßlerstraße 14, D-37073 Göttingen, Germany or e-mail (martin.voracek@univie.ac.at).

theoretically expected and thus meaningful aggregate-level associations of national IQ with numerous other psychological, socioeconomic, and demographic indicators have been obtained. Variables investigated range from atheism (Lynn, Harvey, & Nyborg, 2009), scholastic achievement (Lynn, *et al.*, 2007), fertility (Shatz, 2008), inbreeding depression (Woodley, 2009), health outcomes (Reeve, 2009), and life history traits (Rushton, 2004; Templer, 2008) to homicide (Lester, 2003; Templer, Connelly, Lester, Arikawa, & Mancuso, 2007) and suicide rates (Voracek, 2004, 2005, 2006, 2007a, 2008), to name just a few examples (for a comprehensive review, see Lynn & Vanhanen, 2012b).

One neglected possible correlate of national IQ is the well-known Failed State Index. This is a measure of state vulnerability, to which increasing reference is made by political commentators, economists, and journalists in media reports concerned with present-day geopolitical analyses. Since 2005, this index has been published annually by the U.S. think-tank *Fund for Peace* and the magazine *Foreign Policy*.² Bearing in mind the manifold relationships of national IQ to important socioeconomic and political indicators uncovered in prior research (as partially reviewed above), including evidence for education and intelligence being relevant for countries' political development (Rindermann, 2008), it was thus hypothesized that national IQ might also be meaningfully (i.e., negatively) associated with state vulnerability. To test this hypothesis was the goal of the present research.

METHOD

National IQ estimates were taken from Lynn and Vanhanen (2012a, pp. 19–29). Data on national gross domestic product (GDP) per capita (as of 2012) were taken from the United National Statistics Division website.³ Failed State Index values (as of 2012), available for 177 countries around the world, were gleaned from the *Fund for Peace* website.⁴ The index is compounded of 12 subordinated indicators, of which four are social (indicators i–iv), two economic (indicators v–vi), and six political (indicators vii–xii). In the above order, the social indicators gauged (i) mounting demographic pressures, (ii) massive movement of refugees or internally displaced persons, (iii) legacy of vengeance-seeking group grievance or group paranoia, and (iv) chronic and sustained human flight. The economic indicators assessed (v) uneven economic development along group lines and (vi) sharp or severe economic decline. The political indicators probed (vii) criminalization or delegitimization of the state, (viii) progressive deterioration of public services, (ix) suspension or arbitrary applica-

²<http://global.fundforpeace.org; www.foreignpolicy.com>.

³<http://unstats.un.org/unsd/snaama/Introduction.asp>.

⁴<http://ffp.statesindex.org/rankings-2012-sortable>.

tion of the rule of law and widespread human rights abuse, (x) the security apparatus operating as a "state within the state", (xi) the rise of factionalized elites, and (xii) interventions of other states or external political actors. All 12 indicators were rated from 0.0 (most stable) to 10.0 (least stable), i.e., accurate to the first decimal place and summed to yield the Failed State Index, with a possible range of 0 to 120.

RESULTS

According to the Failed State Index, the top-ten least stable countries in 2012 were Somalia, the Democratic Republic of Congo, Sudan, Chad, Zimbabwe, Afghanistan, Haiti, Yemen, Iraq, and the Central African Republic. The top-ten most stable countries in 2012 were Finland, Sweden, Denmark, Switzerland, Norway, Luxembourg, New Zealand, Ireland, Canada, and Austria. As for further examples, the USA and the UK ranked 19th and 20th, respectively, among the latter group of most stable countries. Across the full set of 177 nations, the Failed State Index showed high, 5-yr. stability, namely 2007 and 2012 values were correlated with Pearson $r = .98$ (two-tailed $p < .001$).

Higher national IQ scores corresponded reliably to higher state stability, i.e., a lower Failed State Index ($r = -.71$, $p < .001$). This was invariably true for all 12 components of the index: correlations were, in the above order of components, $-.77$, $-.61$, $-.38$, $-.68$, $-.66$, $-.70$, $-.62$, $-.75$, $-.56$, $-.62$, $-.54$, and $-.63$ (all $ps < .001$). Additional nonparametric analysis, based on Spearman's rank-order correlation coefficient, showed these results to be robust and not due to outliers or other data peculiarities (details omitted). Controls for the type of national IQ estimate (actually based on study data vs derived from adjacent countries with available data; see Lynn & Vanhanen, 2012a) left the above main finding of higher national IQ corresponding to higher state stability (i.e., lower Failed State Index) essentially unchanged (partial $r = -.69$, $p < .001$). The finding was also preserved with controls for per capita GDP (partial $r = -.54$, $p < .001$).

Of further interest, exploratory factor analysis of the 12 indicators across the 177 countries, using the principal components method, indicated that psychometrically the Failed State Index is a unitary construct. A single factor with an eigenvalue of $\lambda = 9.30$, accounting for 77.5% of the total variance, was extracted, whilst the λ values of all further components were below unity. Loadings on the single factor extracted were, in the above order of components, .80, .73, .66, .61, .78, .71, .88, .84, .81, .88, .83, and .78. The 66 intercorrelations among the 12 indicators ranged from $r = .51$ to .93 (average $r = .75$, all $ps < .001$). Furthermore suggesting unidimensionality of the construct, the internal consistency (Cronbach α) of the 12 indicators was .97. Alternative conceptualizations of internal consistency, such as

McDonald's ω_H (Zinbarg, Revelle, Yovel, & Li, 2005), yielded equally high reliability figures.

DISCUSSION

There are three points of interest in these results. Firstly, the national IQ estimates compiled by Lynn and Vanhanen (2002, 2006, 2012a) appear also useful for macro-level political analyses, as they were meaningfully related to a topical, widely publicized index of state vulnerability. This adds to the credibility and applicability of the data source of national IQ (see Lynn & Vanhanen, 2012b). Secondly, as anticipated, national IQ was indeed inversely related to state vulnerability. The association of these aggregate-level variables was not only reliable, but also strong and preserved with controls. Thirdly, although the index of state vulnerability used here clearly is an elaborated measure, comprised of a total of 12 social, economic, and political facets, this index may be best regarded as a factorially unidimensional construct. This observation potentially has implications for further inquiry along these lines, as it could simplify analyses, in that looking for separate, differentiated effects on the level of the 12 facets of the Failed State Index might be deemed as redundant.

It is certainly true that, although correlation does not mean causality, correlations in general do have underlying causes. However, it is similarly well-known that in many cases the likely causes of correlations arising from aggregate-level, cross-sectional analyses such as the present one remain hard to pinpoint or are even elusive. Hence, the association of national IQ and state vulnerability does not imply causality, which is an important study limitation. It may well be that lower national IQ partly contributes to eventual failing of the state (Rindermann, 2008), but it seems at least equally plausible that the general living conditions in deteriorating states negatively affect national IQ, and perhaps most plausible is that the correlation between these two variables arises because both are only proxies of a suite of even more pervasive historical and societal background variables (e.g., former colonial status, economic underdevelopment, overpopulation, and environmental disadvantages in the broadest sense; see Diamond, 1997, 2005) or evolutionary background variables, such as life history (see Templer, 2008).

It is noteworthy that there is scholarly debate about the validity of national IQ estimates for sub-Saharan countries (for IQ recalculations, see Wicherts, Dolan, & van der Maas, 2010; for counterevidence and a rebuttal, see Lynn & Meisenberg, 2010). However, since sub-Saharan IQs remain the lowest in the nation-sample analyzed here, even when recalculated IQs of Wicherts, *et al.* (2010) are adopted, these differences in IQ estimates virtually have no effect on the present findings.

Despite the above study limitations, the current findings offer various starting points for further tests. The Failed State Index and similar measures may be fruitfully incorporated into cross-national studies of psychometric intelligence, as well as of educational attainment and human development (see Voracek, 2007b). In addition, given the availability of appropriate data, such analyses could also be extended to the intranational level in the case of countries with noticeable regional disparities, conflicts, or upheaval. Furthermore, it would be interesting to investigate time trends in the psychometric structure of the Failed State Index, i.e., whether there are year-to-year changes in its dimensionality or in the differential factor loadings of its 12 facets. Finally, this annually published index would lend itself to usage in cross-lagged analyses of various types, e.g., what short-term effects state deterioration (or, conversely, amelioration) has on educational attainment and public health indicators.

REFERENCES

- BARNETT, S. M., & WILLIAMS, W. (2004) IQ and the wealth of nations. *Contemporary Psychology*, 49, 389-396. [Book review]
- DIAMOND, J. M. (1997) *Guns, germs, and steel: the fates of human societies*. New York: Norton.
- DIAMOND, J. M. (2005) *Collapse: how societies choose to fail or succeed*. New York: Viking.
- DICKERSON, R. E. (2006) Exponential correlation of IQ and the wealth of nations. *Intelligence*, 34, 291-295.
- LESTER, D. (2003) National estimates of IQ and suicide and homicide rates. *Perceptual & Motor Skills*, 97, 206.
- LYNN, R., HARVEY, J., & NYBORG, H. (2009) Average intelligence predicts atheism rates across 137 nations. *Intelligence*, 37, 11-15.
- LYNN, R., & MEISENBERG, G. (2010) The average IQ of sub-Saharan Africans: comments on Wicherts, Dolan, and van der Maas. *Intelligence*, 38, 21-29.
- LYNN, R., MEISENBERG, G., MIKK, J., & WILLIAMS, A. (2007) National IQs predict differences in scholastic achievement in 67 countries. *Journal of Biosocial Science*, 39, 861-874.
- LYNN, R., & VANHANEN, T. (2002) *IQ and the wealth of nations*. Westport, CT: Praeger.
- LYNN, R., & VANHANEN, T. (2006) *IQ and global inequality*. Augusta, GA: Washington Summit.
- LYNN, R., & VANHANEN, T. (2012a) *Intelligence: a unifying construct for the social sciences*. London: Ulster Institute for Social Research.
- LYNN, R., & VANHANEN, T. (2012b) National IQs: a review of their educational, cognitive, economic, political, demographic, sociological, epidemiological, geographic and climatic correlates. *Intelligence*, 40, 226-234.
- REEVE, C. L. (2009) Expanding the g-nexus: further evidence regarding the relations among national IQ, religiosity and national health outcomes. *Intelligence*, 37, 495-505.
- RINDERMAN, H. (2007) The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21, 667-706.

- RINDERMAN, H. (2008) Relevance of education and intelligence for the political development of nations: democracy, rule of law and political liberty. *Intelligence*, 36, 306-322.
- RUSHTON, J. P. (2004) Placing intelligence into an evolutionary framework, or how g fits into the r-K matrix of life-history traits including longevity. *Intelligence*, 32, 320-328.
- SHATZ, S. M. (2008) IQ and fertility: a cross-national study. *Intelligence*, 36, 109-111.
- TEMPLER, D. I. (2008) Correlational and factor analytic support for Rushton's differential K life history theory. *Personality and Individual Differences*, 45, 440-444.
- TEMPLER, D. I., CONNELLY, H. J., LESTER, D., ARIKAWA, H., & MANCUSO, L. (2007) Relationship of IQ to suicide and homicide rate: an international perspective. *Psychological Reports*, 100, 108-112.
- VORACEK, M. (2004) National intelligence and suicide rate: an ecological study of 85 countries. *Personality and Individual Differences*, 37, 543-553.
- VORACEK, M. (2005) National intelligence, suicide rate in the elderly, and a threshold intelligence for suicidality: an ecological study of 48 Eurasian countries. *Journal of Biosocial Science*, 37, 721-740.
- VORACEK, M. (2006) Exponential fitting of suicide rate and national intelligence estimates. *Perceptual & Motor Skills*, 102, 896-898.
- VORACEK, M. (2007a) National differences in intelligence and population rates of suicidal ideation, suicide plans, and attempted suicide. *Perceptual & Motor Skills*, 105, 355-361.
- VORACEK, M. (2007b) Something old, something new, something borrowed and something blue: national IQ and the integration of cognitive-ability research. *European Journal of Personality*, 21, 757-759. [Commentary]
- VORACEK, M. (2008) Cross-national social ecology of intelligence and suicide prevalence: integration, refinement, and update of studies. *Perceptual & Motor Skills*, 106, 550-556.
- WICHERTS, J. M., DOLAN, C. V., & VAN DER MAAS, H. L. J. (2010) A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence*, 38, 1-20.
- WOODLEY, M. A. (2009) Inbreeding depression and IQ in a study of 72 countries. *Intelligence*, 37, 268-276.
- ZINBARG, R. E., REVELLE, W., YOVEL, I., & LI, W. (2005) Cronbach's α , Revelle's β , and McDonald's ω_H : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123-133.

Accepted June 19, 2013.

PERCEIVED LIKELIHOOD OF EXPERIENCING A DESIRABLE VERSUS UNDESIRABLE OUTCOME¹

RON S. GOLD

*Deakin University
Victoria, Australia*

Summary.—In each of 6 studies, some participants were informed that the consumption of certain foods prevents diet-related heart disease; then, they estimated their likelihood of avoiding the disease. Others were informed that failing to consume the same foods contributes to diet-related heart disease; then, they estimated their likelihood of developing the disease. The undesirable outcome was always perceived as less likely than the desirable one. The findings are contrasted with those reported by Bilgin. The difference between the two sets of findings may point to the role event controllability plays in the generation of optimism via motivated reasoning.

People often engage in motivated reasoning: they distort their reasoning in such a way as to favour conclusions they find attractive (Kunda, 1990). From this perspective one would expect people to exhibit optimism about future outcomes. As one manifestation of this, other things being equal, they should estimate their likelihood of experiencing a desirable outcome to be greater than their likelihood of experiencing an undesirable one.

Bilgin (2012, Studies 1 & 2) has reported findings that contradict this expectation. Participants were asked to suppose that, as part of a re-organisation of their workplace, they might be asked to move into an office that was either better (in one condition) or worse (in another condition) than their current office. The changes in the two conditions were identical in qualitative type and magnitude. Estimates of the likelihood of the undesirable outcome were greater than estimates of the likelihood of the desirable one, a result suggestive of pessimism rather than optimism.

This paper too reports data relating to individuals' estimates of the likelihood that they will experience a desirable versus an undesirable outcome. The data represent further analyses from six studies already published (Gold & Martyn, 2004; Gold & Brown, 2009, Studies 1 & 2; Gold & Nguyen, 2010; Gold & de Sousa, 2012, Studies 1 & 2). In each study, participants were informed of a link between diet when one is young and the health of one's heart in later life. Exactly the same information was framed

¹Address correspondence to Assoc. Prof. Ron S. Gold, School of Psychology, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia or e-mail (ron.gold@deakin.edu.au).

differently in different conditions. In one condition, the focus was on consumption of certain foods that prevent diet-related heart disease and participants estimated their likelihood of avoiding the disease; in another condition, the focus was on failure to consume the same foods and participants estimated their likelihood of developing the disease. Absolute rather than comparative likelihood was assessed; a 7-point Likert-type scale, with a higher score indicating greater likelihood, was employed. Each sample comprised students attending Deakin University. Full details of methodology are in the published papers. Table 1 presents the means and standard deviations of participants' estimates that they would develop/avoid the disease, together with the results of a test of the difference between these estimates. In each case, there was a large effect such that the undesirable outcome was perceived as less likely than the desirable one.

TABLE 1
COMPARISON OF MEAN LIKELIHOOD ESTIMATES FOR UNDESIRABLE AND DESIRABLE OUTCOMES

Study	Undesirable Outcome		Desirable Outcome		Comparison of Means	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Gold & Martyn (2004)	2.98	1.16	4.27	1.30	$t_{173} = 6.95\ddagger$	1.05
Gold & Brown (2009, Study 1)	2.88	1.20	4.28	1.34	$t_{113} = 5.91\ddagger$	1.10
Gold & Brown (2009, Study 2)	2.80	1.06	3.95	1.28	$t_{137} = 5.69\ddagger$	0.98
Gold & Nguyen (2010)	3.25	1.12	4.11	1.56	$t_{111} = 3.35\ddagger$	1.07
Gold & de Sousa (2012, Study 1)	2.99	1.26	4.52	1.43	$t_{157} = 7.17\ddagger$	1.14
Gold & de Sousa (2012, Study 2)	3.22	1.15	4.56	1.20	$t_{88} = 5.40\ddagger$	1.14

$\ddagger p < .001$.

Both of Bilgin's (2012) studies concerned the 'moving office' scenario. All our studies concerned the same threat to heart health. Student samples were used in both Bilgin's studies and ours. These limitations mean that the generalizability of the findings to other events and samples remains to be established. Nevertheless, to the author's knowledge, Bilgin's and our studies are the only ones in which desirable and undesirable outcomes were so presented that their content was very closely matched. The contrast between the two sets of findings is therefore striking.

While Bilgin's (2012) scenarios and ours clearly differed in numerous respects, it may be suggested—consistent with Bilgin's own comments—that the difference in findings may have to do with the *controllability* of the events whose perceived likelihood was assessed. In Bilgin's studies, the move to a new office was presented as outside the participants' control,

while in the present studies the health outcome was presented as controllable through diet. Use of an uncontrollable event in Bilgin's studies may have prevented the use of motivated reasoning. Kunda (1990) argued that motivated reasoning is constrained by a need to remain in plausible accord with reality. For a controllable event there are, by definition, behaviours that could change the likelihood of the outcome. This makes for a convenient route for motivated reasoning to occur: individuals can often find at least some favourable (unfavourable) behaviours that they plausibly engage in (avoid) and selectively focus on these to arrive at a congenial conclusion. But this strategy is not available for an uncontrollable event. The difference between the two sets of findings may thus point to the importance of controllability in the generation of optimism via motivated reasoning.

REFERENCES

- BILGIN, B. (2012) Losses loom more likely than gains: propensity to imagine losses increases their subjective probability. *Organizational Behavior and Human Decision Processes*, 118, 203-215.
- GOLD, R. S., & BROWN, M. G. (2009) Explaining the effect of event valence on unrealistic optimism. *Psychology, Health & Medicine*, 14, 262-272.
- GOLD, R. S., & DE SOUSA, P. N. (2012) When does event valence affect unrealistic optimism? *Psychology, Health & Medicine*, 17, 105-115.
- GOLD, R. S., & MARTYN, K. (2004) Event valence and unrealistic optimism: further evidence. *Psychological Reports*, 95, 464-466.
- GOLD, R. S., & NGUYEN, H. (2010) Event valence and estimates of own, average person's, and randomly chosen person's risk. *Psychological Reports*, 107, 682-684.
- KUNDA, Z. (1990) The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.

Accepted August 16, 2013.

OBJECT EXPLORATION IN 3- TO 7-YEAR-OLD CHILDREN¹

MAGDALENA ŚNIEGULSKA

Warsaw School of Social Sciences and Humanities

WOJCIECH PISULA

*Institute of Psychology
Polish Academy of Sciences*

Summary.—The aim of the present study was to analyze free exploration of an unfamiliar, novel object in a semi-natural environment in 189 preschool children aged 3 to 7 years. The study used a purpose-built, novel, and complex stimulus object. Each child was shown the complex object lying on the floor and asked to explore it for 15 minutes. The child's behavior was video recorded. Age-dependent differences were found in the children's exploratory activity. No sex differences were found. Cluster analysis showed four main behavioral patterns that differed in their qualitative and quantitative parameters. Cluster membership was mainly age-dependent and not sex-specific. The results showed individual and developmental differences among the preschoolers with respect to their exploratory activity. This should be taken into account by school psychologists and teachers when designing the educational tasks and play situations.

Various forms of exploration belong to the basic components of the human and animal behavioral repertoire. For the members of long-living species, whose development is mostly or primarily learning-based, exploration of the environment is a constant task. There are reasons to believe that exploratory behaviors are the basis of cognitive development in humans and animals (Rabinowitz, Moely, Finkel, & McClinton, 1975; Pisula, 1998, 2009). The importance of object use by children has been shown recently by Pellegrini and Hou (2011). The role of qualitative properties and complexity of the objects has been demonstrated by Schulz and Bonawitz (2007) and Switzky, Haywood, and Isett (1974). It seems that the informative value of the objects studied plays a crucial part of the child exploratory behavior. In the current study, an essential aspect of the test object was its novelty.

In most mammals, exploratory activity peaks after weaning time (Renner & Pierre, 1998; Pisula, 2008). This period is associated with profound changes in the nervous system, as reported in both animal (Byers & Walker, 1995) and human studies (Casey, Tottenham, Liston, & Durston,

¹Address correspondence concerning this paper to Wojciech Pisula, Institute of Psychology, PAS, Warsaw or e-mail (wojciech.pisula@wp.pl). This research was supported by the research project UMO – 2011/01/B/HS6/02458.

2005). In humans, the weaning time varies depending on cultural factors. The optimum age for weaning in human children is thought to be 2.5 years, although in traditional hunter-gatherer societies it may be extended to approximately 4 years (Kennedy, 2005). As far as brain development is concerned, the brain of a 5- to 7-year-old child is 90% of weight of an adult brain (Dekaban & Sadowsky, 1978), and its development continues for many years (Lenroot & Giedd, 2006). Some authors have reported sex differences in the development of the nervous system (Ellison & Nelson, 2009). Around 5 to 7 years of age, learning assumes the form of regular school education, which is an important developmental milestone. It appears that the period from around 3 years of age to 6 or 7 years of age is a vital stage during which the framework of information-seeking strategies emerges.

Research on exploration in children is strongly influenced by the approach founded on object exploration analysis. Studies using a novel object are sometimes referred to as "object play" instead of "object exploration" (see Vig, 2007). In terms of Berlyne's classification (1963), these manipulatory and investigatory responses fall within specific exploration categories, rather than diverse exploration. The main body of the literature has focused on exploration in infants (see Bourgeois, Khawar, Neal, & Lockman, 2005). The relatively few reports on exploration at later stages of life show that over the course of development the rate of specific (object-oriented) exploration tends to increase relative to diverse exploration (Vlietstra, 1978; Pisula, 1998). It seems that exploratory behavior concentrated on specific objects is particularly important for the development of cognitive function, which is essential for the mastery of the educational tasks the child faces at school (Vig, 2007) and plays a key role in learning about causality (Schulz, Standing, & Bonawitz, 2008).

The aim of the present study was to analyze free exploration of an unfamiliar, novel object in a semi-natural environment in children of 3 to 7 years. The study followed an etho-experimental approach to create a detailed catalogue of observed behaviors based on ethograms generated for animal behavior. Identification of various forms of responding to novel objects in highly structured situations in children at the age when they start school could have significant implications for planning school activities. These data may enrich the stimulation of children's exploratory activity by manipulating the novelty and complexity of the stimuli.

METHOD

Participants

Preschool children aged 37 to 90 months ($N = 189$; $M = 63.3$; $SD = 12.9$) participated in the study. The study was conducted in private and public kindergartens in Poland: in Warsaw (the capital, population 1,708,000),

two cities with populations under 50,000,*and in a small village (population < 200). The study group's demographic details are presented in Table 1. The distribution of age categories reflects the organization of the pre-school groups available for the study, and the attempt to create subsamples of approximately equal sizes.

TABLE 1
SAMPLE CHARACTERISTICS BY GENDER AND PLACE OF RESIDENCE AND BY GENDER AND AGE GROUP

Residence		Boys		Girls		
Warsaw		28		18		
Mińsk Mazowiecki		20		23		
Lesko		16		35		
Łbiska		22		27		
Age Range	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
37–54 months	25	48.08	4.25	31	48.00	4.79
55–66 months	29	61.48	3.63	26	60.12	3.95
67–90 months	32	75.69	5.93	46	76.17	6.82
Overall	86	62.86	12.22	103	63.64	13.40

The study used a complex rectangular object (50 × 45 × 10 cm). The object was made of wood and upholstered with a multi-colored material. It consisted of 16 parts. Some parts resembled human figures; others referred to well-known and popular objects used in play and were made of wood, fabric, and plastic. Although parts of the object set may look familiar, it has been shown in the pilot study that as a whole they constituted an unfamiliar stimulus object for the children. These objects were: a goat, a frog, ball-shaped bells, puzzle pieces, a doll, a green bag, cogs, a pouch with a zipper, marbles, an elephant-shaped block, a fish-shaped block, a little man, a padlock, keys, and a fish (Fig. 1). The object was designed to contain elements attractive to boys and girls alike. Some objects were available for direct observation while others were placed inside zipped pouches or multicolored bags. Some elements were detachable, while others were fixed permanently to the stimulus object. The pilot sessions showed that children did not recognize the object as familiar.

Procedure

The measurements were carried out in the classrooms. Prior to each measurement, the four female experimenters were required to get to know the children and spend approximately one hour with them, at a similar time of day (after compulsory classes, during the so-called free time). The



FIG. 1. Stimulus object used to elicit exploration in children.

study was carried out on an individual basis, in a room provided by the kindergarten in the same classroom the child attended. There were two experimenters in the room. One experimenter recorded the experiment with a video camera. Each child was shown the complex object lying on the floor and asked to explore it. The study enrolled only those volunteering children who agreed to take part and whose parents provided a written consent for their participation and video recording. A 15-min. exploration session was recorded. Since many of the children stopped their activity before the 15-min. period had elapsed, only the first 10 min. of each session were subsequently analyzed.

Parents were informed in detail about the purpose and method of the study and gave their consent for their children's participation. The study conformed with research ethics regulations in Poland.

Coding Children's Behaviors

Before proper data analysis, three competent raters generated a list of 12 measurable behaviors on the basis of the following procedure. The first stage involved analysis of five randomly selected recordings by three child psychologists. The initial list of 21 measurable behaviors was then discussed and re-analyzed. In the next stage, the three observers grouped the measured behavior, reducing the list to 12 by consensus. The list of coded behaviors with descriptions is provided in Table 2.

All identified behaviors were coded using the EthoLog software, Version 2.2 (Ottoni, 2000). Frequency and duration were assessed for each behavior. In addition, the program makes it possible to trace transitions between individual behaviors. The number of the object's elements

TABLE 2
CODED BEHAVIORAL CATEGORIES GENERATED FROM OBSERVED BEHAVIORS

Behavior Label	Code	Description
Pausing	A	The child sits or stands next to the object, not looking at the object or the experimenter, not interacting with the object.
Refusal	B	The child clearly refuses to continue taking part in the test (e.g., "I already know everything," "I'm done," "I want to get out," etc.).
Manipulation	C	Any interaction with the object and its elements involving touching but not detaching/restoring previously detached elements.
Two activities simultaneously—commenting on his/her own activity	D	The child comments on his/her activity (talks to himself/herself or the experimenter while exploring the object).
Dismounting/detaching elements of the object	E	The child detaches an element from the object.
Staring	F	The child sits or stands next to the object and looks at it without touching.
Object play	G	Using the properties of the explored element, usually according to its intended purpose, behavior demonstrated as if "in the company" of the object, qualitatively different from object exploration (e.g., the child plays with a doll, talks to it, gives orders such as: "and now you must eat the fish," etc.).
Eye contact with the experimenter	H	The child sits or stands looking at the experimenter, not performing any other activities.
Restoring	I	The child restores previously detached elements of the object back to their original place.
Soliloquizing	J	The child talks to himself/herself without looking at the experimenter or performing any other activities.
Talking to the experimenter	K	The child talks to the experimenter, clearly addresses him/her, looks at the experimenter, asks him/her questions, tells stories, describes elements while looking at the experimenter and not performing any other activities.
Simultaneous use of two hands and two objects separately	L	Using two hands to explore two separate elements of the object at the same time.

explored during each measurement session and time spent on exploring the whole object were also calculated.

RESULTS

To assess whether the overall time spent by each child on active exploration of the stimulus object (time spent on behavioral categories: C, E, G, I, L shown in Table 2) differed by age or sex, a two-way analysis

TABLE 3
RESULTS OF BETWEEN SUBJECTS ANALYSIS OF VARIANCE BY AGE GROUP

Age Group	Example	M	SD
Active object exploration time (sec.)			
37-54 mo.	56	125.44	154.23
55-66 mo.	55	203.79	160.86
67-90 mo.	78	235.41	166.07
All	189	193.63	166.83
$F_{2,188} = 7.75, p < .001$			
Number of stimulus object elements explored			
37-54 mo.	56	5.00	5.95
55-66 mo.	55	7.76	6.15
67-90 mo.	78	8.76	6.42
All	189	6.22	6.38
$F_{2,188} = 6.22, p < .01$			

of variance (ANOVA) was conducted. It revealed statistically significant differences in terms of the overall time spent on exploring the stimulus object, which are shown in detail in Table 3 (part A). The *post hoc* tests showed the youngest children spent less time in object exploration than the remaining two groups. No sex differences or interaction of age and sex were noted.

To investigate the relation between exploration outcome (measured by the number of elements explored) and the child's age and sex, an ANOVA was conducted. There were statistically significant differences in terms of the number of elements explored, which are shown in detail in Table 3, part B. The *post hoc* tests showed the youngest children explored fewer objects than the remaining two groups. No sex differences were found.

An attempt was also made to identify relatively homogeneous groups with similar exploration profiles through cluster analysis using the K-means algorithm. The aim of this analysis was to identify consistent behavior patterns demonstrated by children during the study. The analysis included variables describing behavioral sequences obtained by combining variables describing the 12 basic identified behaviors. Thus, if behavior A was followed by behavior B, the sequence was described as variable AB. The cluster analysis included 144 variables describing behavioral transitions. The four clusters obtained are shown in Table 4. Diagrams describing behavior patterns were plotted for individual clusters. They are shown in Figure 2.

No relation was found between membership in a given cluster and the child's sex or the population of the city where the experiment was

TABLE 4
CLUSTER CHARACTERISTICS

Cluster	Cluster Size	Within Sum of Squares
1	65	11449.77
2	12	5506.41
3	78	6675.38
4	34	5179.99

conducted. However, the clusters differed by the age of the children in each cluster and the number of elements explored by the child, as well as the time spent exploring the object. These differences are presented in Table 5.

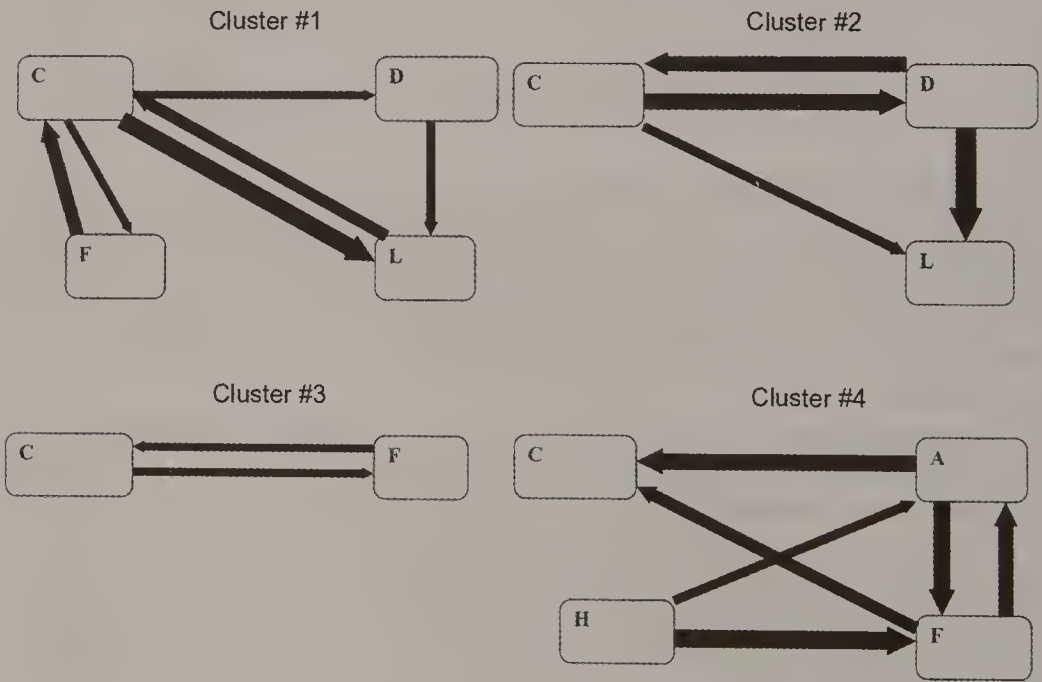


FIG. 2. Behavioral patterns shown by children that formed four clusters. The thickness of the arrows reflect relative frequency of the given behavioral act occurrence. See Table 2 for the description of symbols in the boxes.

DISCUSSION

Exploratory behavior is known to change over the course of children's development. Initially passive, with time they become increasingly active and systematic, incorporating previous knowledge (Piaget & Inhelder, 1956). This trend was confirmed in the present study. Older children scored higher on exploration, as measured by the number of elements of the object they explored. This effect was independent of the child's sex and

TABLE 5
CLUSTER MEMBERSHIP, CHARACTERISTICS OF EXPLORATION, AND AGE IN CHILDREN

Cluster	Age (mo.)		Overall Time Spent on Exploring the Stimulus-object		Investigated Elements of the Stimulus-object	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	68.54	11.70	568.58	74.57	11.38	3.29
2	62.25	14.21	572.17	52.38	10.25	4.58
3	61.00	12.62	322.14	258.35	3.04	4.00
4	58.76	11.95	584.35	51.79	3.65	4.60
$F_{3,188} = 63.81, p < .001$			Kruskal-Wallis $H = 51.2, p < .001$		$F_{3,188} = 6.41, p < .001$	

individual experience, because the object was novel for all participants. Thus, the results clearly reflect the developmental character of restructuring in exploratory behaviors of children within the age range of the sample. Large individual differences within the age groups underscore the need to offer individualized learning environments for children. This result is consistent with earlier studies (Abravanel, 1968; Vandenberg, 1984).

No sex differences were observed in the present study. The presence of such differences is often disputed. Few authors have reported differences in object manipulation between boys and girls under two years of age. Most studies have concluded that there are no sex differences at this stage of life (Yarrow, McQuiston, MacTurk, McCarthy, Klein, & Vietze, 1983; van den Boom, 1994). The results of research on exploration in older children are less consistent. Some suggest the presence of sex differences in terms of greater vigor and diversity in the object manipulation of boys, along with greater success in haptic exploration tasks in girls (Hutt, 1970; Daldry & Russell, 1982). Still, most authors report no sex differences (McLloyd & Ratner, 1983; Schneider, Moch, Sandfort, Auerswald, & Walther-Weckman, 1983).

Cluster and sequential analysis of exploratory behavior yielded four major exploration patterns. Clusters 1 and 2 were characteristic of older children and were associated with exploration of a large number of elements. Clusters 3 and 4, typically demonstrated by younger children, were characterized by fewer elements explored. The first pattern, observed primarily in the oldest children (M age = 68.5 mo.), was characterized by the presence of behaviors such as manipulation, parallel manipulation (simultaneously using two hands to explore two objects separately), quiet visual examination of the object, and performing two actions at the same time. This type of behavior is consistent with Piagetian "systematic exploration" (Piaget & Inhelder, 1956), which proceeds in accordance with a

plan. Children in this group used manipulation, exploring the properties of multiple elements at once (parallel manipulation) and, after a while, focused their attention on one element of the object (manipulation). Once a child established knowledge of a given element, he or she started looking for another stimulus. At the same time, the children asked questions and offered comments (two actions at the same time), perhaps to collect additional information or to structure their knowledge. Manipulation was often preceded by staring at the object, perhaps with the purpose of analyzing and planning further exploration. The time spent examining the object of exploration by the children in this group was the most effective in terms of the mean number of elements investigated, which was the highest of all groups.

Cluster 2 behavior was typical for slightly younger children (M age = 62.2 mo.). The dominant behaviors in this group were manipulation, parallel manipulation, and performing two actions simultaneously (manipulating with the object and commenting on his or her own activity). The main activity focused on manipulation and two simultaneous actions: the child explored an element of the object and commented on his/her activity. Children in this group also used manipulation, explored the properties of multiple elements at once (parallel manipulation), and, after a while, focused their attention on a selected element of the object (manipulation). Once it seemed familiar, they would look for another stimulus. Lack of staring, perhaps indicating less analysis and planning, seems to be responsible for the less efficient exploration compared to Cluster 1. Children spent a little more time examining the object compared with the previous group, but their efficiency was lower.

Another pattern of exploratory behavior characteristic of yet younger children (Cluster 3, M age = 61.0 mo.) was associated with the presence of manipulation alternating with staring. This pattern may be consistent with Piagetian non-systematic active exploration (Piaget & Inhelder, 1956). Children in this group devoted the least time to the object, and the efficiency of their exploration was the lowest. For them, familiarization with the investigated object required more time and resources.

Finally, in Cluster 4 there was a pattern of exploration demonstrated by the youngest children (M age = 58.8 mo.), who according to Piaget's classification (Piaget & Inhelder, 1956) were in the stage of passive exploration. Interestingly, this behavior pattern was associated with the longest time devoted to the object, coinciding with a relatively low efficiency of exploration. The pattern featured four types of behavior: eye contact with the experimenter, staring at the object, pausing, and manipulation. The main sequence of behaviors proceeded as follows: the child established eye contact with experimenter or froze (no activity) or stared at the object. Star-

ing was often not followed by further exploration. These types of activities were dominant, taking up most of the time spent exploring the object.

A number of investigators have emphasized the importance of children's cognitive assessment, e.g., for estimating the effectiveness of interventions or predicting the child's functioning at school (Borden, Brown, Wynne, & Schleser, 1987). Children at later stages of cognitive development are more likely to use problem-solving strategies consistently and to employ recursive thought processes (Forehand & Wierson, 1993). Researchers investigating exploratory behavior often point out the relation between sustained attention during object manipulation and developmental assessments (Kagan, Lapidus, & Moore, 1978; Caruso, 1993). Behavior during exploration also has been found to be associated with psychological adjustment (McReynolds, Acker, & Pietila, 1961). Since there are still no simple and easy-to-use cognitive assessment methods, research on exploration in children seems all the more important.

Between 3 and 7 years of age, depending on the particular educational system, children commence systematic school education. This is a crucial period of rapid cognitive changes, along with wide individual differences in development and ability to adapt to novelty. The popular trend of lowering the school age in Poland and many other countries should be preceded by detailed assessment of skills and differences between children in this age group in terms of their ability to cope with novel stimulation.

REFERENCES

- ABRAVANEL, E. (1968) The development of intersensory patterning with regard to selected spatial dimensions. *Monographs of the Society for Research in Child Development*, 33(2, Serial No. 118), 1-52.
- BERLYNE, D. E. (1963) Motivation problems raised by exploratory and epistemic behavior. In S. Koch (Ed.), *Psychology: a study of science*. Vol. 5. New York: McGraw-Hill. Pp. 284-364.
- BORDEN, K. A., BROWN, R. T., WYNNE, M. E., & SCHLESER, R. (1987) Piagetian conservation and response to cognitive therapy in attention deficit disordered children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 28, 755-764.
- BOURGEOIS, K. S., KHAWAR, A. W., NEAL, A. S., & LOCKMAN, J. J. (2005) Infant manual exploration of objects, surfaces, and their interrelations. *Infancy*, 8(3), 233-252.
- BYERS, J., & WALKER, C. (1995) Refining the motor training hypothesis for the evolution of play. *American Naturalist*, 146, 25-40.
- CARUSO, D. A. (1993) Dimensions of quality in infant's exploratory behavior: relationships to problem-solving ability. *Infant Behavior and Development*, 16, 441-454.
- CASEY, B. J., TOTTENHAM, N., LISTON, C., & DURSTON, S. (2005) Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitive Sciences*, 9(3), 104-110.
- DALDRY, A. D., & RUSSELL, P. A. (1982) Sex differences in the behavior of preschool children with novel and familiar toys. *Journal of Genetic Psychology*, 141, 3-6.

- DEKABAN, A. S., & SADOWSKY, D. (1978) Changes in brain weight during the span of human life: relation of brain weights to body heights and body weights. *Annals of Neurology*, 4, 345-356.
- ELLISON, P., & NELSON, A. (2009) Brain development: evidence of gender differences. In E. Fletcher-Janzen (Ed.), *The neuropsychology of women*. New York: Springer Science. Pp. 11-30.
- FOREHAND, R., & WIERSON, M. (1993) The role of developmental factors in planning behavioral interventions for children: disruptive behavior as an example. *Behavior Therapy*, 24, 711-724.
- HUTT, C. (1970) Curiosity in young children. *Science Journal*, 6, 68-72.
- KAGAN, J., LAPIDUS, D. R., & MOORE, M. (1978) Infant antecedents of cognitive functioning: a longitudinal study. *Child Development*, 49, 1005-1023.
- KENNEDY, G. E. (2005) From the ape's dilemma to the weanling's dilemma: early weaning and its evolutionary context. *Journal of Human Evolution*, 48, 123-145.
- LENROOT, R. K., & GIEDD, J. N. (2006) Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*, 30, 718-729.
- MCLLOYD, V. C., & RATNER, H. H. (1983) The effects of sex and toy characteristics on exploration in preschool children. *Journal of Genetic Psychology*, 142, 213-224.
- MCREYNOLDS, P., ACKER, M., & PIETILA, C. (1961) Relation of object curiosity to psychological adjustment in children. *Child Development*, 32, 393-400.
- OTTONI, E. B. (2000) EthoLog 2.2: a tool for the transcription and timing of behavior observation sessions. *Behavior Research Methods, Instruments, & Computers*, 32(3), 446-449.
- PELLEGRINI, A. D., & HOU, Y. (2011) The development of preschool children's (*Homo sapiens*) uses of objects and their role in peer group centrality. *Journal of Comparative Psychology*, 125(2), 239-245. DOI: 10.1037/a0023046
- PIAGET, J., & INHELDER, B. (1956) *The child's conception of space*. London, UK: Routledge & Kegan Paul.
- PISULA, W. (1998) Integrative levels in comparative psychology—the example of exploratory behavior. *European Psychologist*, 3, 62-69.
- PISULA, W. (2008) Play and exploration in animals—a comparative analysis. *Polish Psychological Bulletin*, 39(2), 104-107.
- PISULA, W. (2009) *Curiosity and information seeking in animal and human behavior*. Boca Raton, FL: BrownWalker.
- RABINOWITZ, F. M., MOELY, B. E., FINKEL, N., & MCCLINTON, S. (1975) Effects of toy novelty and social interaction on the exploratory behavior of preschool children. *Child Development*, 46, 286-289.
- RENNER, M. J., & PIERRE, P. J. (1998) Development of exploration and investigation in the Norway rat (*Rattus norvegicus*). *The Journal of General Psychology*, 125(3), 270-291.
- SCHNEIDER, K., MOCH, M., SANDFORT, R., AUERSWALD, M., & WALTHER-WECKMAN, K. (1983) Exploring a novel object by preschool children: a sequential analysis of perceptual, manipulatory, and verbal exploration. *International Journal of Behavioral Development*, 6, 477-496.
- SCHULZ, L. E., & BONAWITZ, E. B. (2007) Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, 43(4), 1045-1050. DOI: 10.1037/0012-1649.43.4.1045

- SCHULZ, L. E., STANDING, H. R., & BONAWITZ, E. B. (2008) Word, thought, and deed: the role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology*, 44(5), 1266-1276.
- SWITZKY, H. N., HAYWOOD, H. C., & ISETT, R. (1974) Exploration, curiosity, and play in young children: effects of stimulus complexity. *Developmental Psychology*, 10(3), 321-329.
- VAN DEN BOOM, D. C. (1994) The influence of temperament and mothering on attachment and exploration: an experimental manipulation of sensitive responsiveness among lower class mothers with irritable infants. *Child Development*, 65, 1457-1477.
- VANDENBERG, B. (1984) Developmental features of exploration. *Developmental Psychology*, 20, 3-8.
- VIG, S. (2007) Young children's object play: a window on development. *Journal of Developmental and Physical Disabilities*, 19, 201-215.
- VLIETSTRA, A. G. (1978) Exploration and play in preschool children and young adults. *Child Development*, 49, 235-238.
- YARROW, L. J., MCQUISTON, S., MACTURK, R. H., MCCARTHY, M. E., KLEIN, R. P., & VIETZE, P. M. (1983) Assessment of mastery motivation during the first year of life: contemporaneous and cross-age relationships. *Developmental Psychology*, 19, 159-171.

Accepted August 22, 2013.

EFFECTS OF PROLONGED WAKEFULNESS: THE ROLE OF PERIOD3 GENOTYPES AND PERSONALITY TRAITS¹

GIUSEPPE BARBATO, ANTONIO COSTANZO, CIRO DELLA MONICA

Department of Psychology
Second University of Naples, Italy

PAOLO D'ONOFRIO

Stress Research Institute
University of Stockholm, Sweden

FLAVIA CERRATO

Department of Environmental Science
Second University of Naples, Italy

VITTORIA DE PADOVA

Department of Psychology
Second University of Naples, Italy

Summary.—The roles of personality traits, as assessed by Eysenck Personality Inventory, and of the clock gene PERIOD3 (PER3) were analysed on the subjective effects of prolonged wakefulness. A sample of 70 healthy participants (7 men, 63 women; M age = 24.2 yr., SD = 3.2) was studied during forced wakefulness between 7:30 p.m. and 9:30 a.m. According to Eysenck's arousal model, it was hypothesized that prolonged wakefulness might affect in a different way those classified as Introverted and Extraverted. During the forced wakefulness period, the Introverted group showed greater decrease in subjective measures of vigilance than did the Extraverted group, but no differences were observed between groups with high and low scores on Psychoticism and Neuroticism. Prolonged wakefulness had a negative effect on subjective sleepiness and mood in all three PER3 polymorphisms analysed.

Sleep deprivation is a sleep–wake manipulation used to test sleep functions. Deprivation affects different neuro-cognitive systems, producing impairments on vigilance, working memory, and executive functions (Dinges & Kribbs, 1991; Harrison & Horne, 1998). Although agreement exists in literature on negative consequences of sleep deprivation, inter-individual differences have been observed, with some individuals showing great sensitivity to sleep loss while others are less affected (Van Dongen, Baynard, Maislin, & Dinges, 2004; Caldwell, Mu, Smith, Mishory, Caldwell, Peters, *et al.*, 2005; Van Dongen, Vitellaro, & Dinges, 2005).

Personality traits and genetic factors have been suggested to explain such differences. The seminal work of Corcoran (1962, 1972) showed that during sleep deprivation extraverts had a progressive deterioration on

¹Address correspondence to Dr. Giuseppe Barbato, Department of Psychology, Second University of Naples, Viale Ellittico 31, 81100, Caserta, Italy or e-mail (Giuseppe.Barbato@unina2.it).

“tracking” performance, whereas introverts’ performance improved with increasing deprivation. Also, Taylor and McFatter (2003) found that after sleep deprivation extraverts performed worse than introverts on cognitive tasks, and Killgore, Richards, Killgore, Kamimori, and Balkin (2007) reported that higher scores on extraversion at baseline were associated with greater decline in alertness and psychomotor vigilance during the first 26 hours of a prolonged sleep deprivation.

According to Eysenck’s model of personality, the respectively high and low arousal in introverts and extraverts can explain the different behavior of these two traits under sleep-deprivation conditions (Matthews, 1992). In recent years, cognitive performance in response to sleep restriction has been associated with polymorphisms of the “clock gene” PERIOD3 (PER3): compared to those with PER3 4/4 genotypes, those with PER3 5/5 show a greater cognitive decline after total sleep deprivation (Viola, Archer, James, Groeger, Lo, Skene, *et al.*, 2007; Dijk & Archer, 2010). Goel, Banks, Mignot, and Dinges (2009), however, did not find differences between PER3 4/4, PER3 4/5, and PER3 5/5 genotypes in cognitive performance and sleepiness across five consecutive nights of sleep restriction.

Subjective self-perception and attitude toward the stressful condition of sleep deprivation could also affect performance, since after sleep deprivation participants can show acute worsening of mood, irritability, and decreased motivation (Johnson & MacLeod, 1973; Heiser, Dickhaus, Oppen, Hemmeter, Remschmidt, Wesemann, *et al.*, 2001; Durmer & Dinges, 2005). The present study’s purpose is to further understand the factors involved in individual responses to sleep deprivation, looking at effects on vigilance, activation, and mood. Considering results from previous studies:

Hypothesis. Better tolerance of a sleep deprivation condition is expected in those characterized by introversion and/or having the PER3 4/4 allele.

METHOD

Participants

Healthy Caucasian volunteers ($N = 70$; 7 men, 63 women; M age = 24.2 yr., $SD = 3.19$) were recruited from undergraduate psychology classes at the Second University of Naples. The study was conducted in accordance with the ethical standards of the Declaration of Helsinki. All participants were informed about the nature and procedures of the study and signed a written consent form. To exclude individuals with relevant sleep impairments, use of psychotropic drugs or substance abuse, a questionnaire assessing life and sleep habits was administered. Participants had no his-

tory of psychiatric illnesses (DSM-IV, Axis I), and were not affected by any significant medical or neurological illness. Participants were instructed to maintain a regular sleep-wake cycle during the week which preceded the experimental night session. Compliance with the schedule was assessed by self-report sleep diaries (Monk, Reynolds, Kupfer, Buysse, Coble, Hayes, *et al.*, 1994).

Procedure

Participants arrived in the laboratory at 7:30 p.m. and left at 9:30 a.m. the next day; they underwent a period of prolonged wakefulness, which started at 9:00 p.m. and ended at 8:00 a.m. the next day. During the experimental night session, groups of randomly assigned participants (between 4 and 6 for each session) were continuously monitored and kept engaged by two laboratory technicians in a common living area. Participants underwent the experimental night session in a double blind condition for both personality trait assignment and genotype. Indoor light was of 150 lux at the level of the participants' eyes. Neither coffee nor alcoholic beverages were allowed. Sleepiness was assessed each hour (from 9:00 p.m. to 8:00 a.m.).

Measures

Sleepiness.—The Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990), a 9-point Likert-type scale with anchors 1: Extremely alert and 9: Extremely sleepy, fighting sleep, was administered each hour from 9:00 p.m. to 8:00 a.m. The KSS is widely used to assess subjective sleepiness, and the scale's points have shown a close relationship to behavioral and EEG indicators of sleepiness (Åkerstedt & Gillberg, 1990; Kaida, Takahashi, Åkerstedt, Nakata, Otsuka, Haratani, *et al.*, 2006).

Subjective vigilance, activation, and mood were assessed by visual analogue scales (VAS; Monk & Embrey, 1989). The scales were 100-mm horizontal lines, anchored by word descriptors at each end. The participant was asked to mark the point on the line which best described his or her current state between the extremes of a continuum from 0: Very low to 10: Very high (Monk & Embrey, 1989). VAS scales were administered each hour from 9:00 p.m. to 8:00 a.m.

Circadian preferences were assessed by the Italian version of the Morningness-Eveningness questionnaire reduced form (r-MEQ; Adan & Almirall, 1992; Natale, 1999). The validity of the r-MEQ has been assessed by Natale (1999). The translation has internal consistency reliability (Cronbach $\alpha = .71$) similar to the original form. The correlation between r-MEQ and MEQ total scores was strong ($r = .90, p < .001$).

Personality.—To evaluate differences in personality traits, the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1964) was administered.

All individuals were characterized in dichotomized groups according to their scores on the Eysenck traits. Groups scoring high and low on personality traits comprised, respectively, those who scored at one standard deviation above the total group mean (High) and one standard deviation below the total group mean (Low), according to Taylor and McFatter (2003). The Low Extraversion group is described here as the Introverted group, and the High Extraversion group is described as the Extraverted group.

PERIOD3 Genotyping

Genomic DNA was extracted from buccal cells using the Gentra Pure-gene buccal cell kit (by Qiagen) and polimerase chain reaction (PCR) assay was performed to measure the length polymorphism of 54 base pairs (bp) in exon 18 of the *Per3* gene. The PCR primers used were: forward, 5'-CAAAATTTTATGACACTACCAGAATGGCTGAC-3', and reverse, 5'-AACCTTGTACTTCCACATCAGTGCCTGG-3' (Ebisawa, Uchiyama, Kajimura, Mishima, Kamei, Katoh, *et al.*, 2001). PCR was performed in a 50 μ L volume made up of 5 μ L (10 ng/ μ L) of genomic DNA, 0.8 μ M of each primer, 0.2 mM dNTP, 5 μ L 10 \times buffer [160 mM (NH₄)₂SO₄, 670 mM Tris-HCl pH 8.8, 0.1% stabilizer], 1.5 mM MgCl₂, and 2 units of Biotaq Red DNA polymerase (Bio-line). A positive DNA control and a negative water control were included with each PCR plate of samples. The PCR cycling conditions were 3 min. at 94°C, followed by 35 cycles of 45 sec. at 94°C, 45 sec. at 58°C, and 45 sec. at 72°C, with a final step at 72°C for 3 min. to allow for the complete extension of all PCR fragments.

The products were resolved and separated for 30 min. at 150 V on a 1.5% agarose gel stained with ethidium bromide. After electrophoresis, homozygous alleles with 5 repeats were represented by a DNA band with size at 635 bp. Whereas homozygous alleles with 4 repeats were represented by a DNA band with size at 581 bp, heterozygotes displayed a combination of both alleles (635 and 581 bp). Approximately 5% of the samples were duplicated to assure quality control in genotyping; two reviewers separately did genotype scoring to confirm results.

Analysis

Following the sleep deprivation sessions, all participants were assigned to personality and polymorphism groups according to the procedures previously described. Statistical analysis for *PER3* was conducted through a 3 \times 12 mixed-model repeated-measures analysis of variance (ANOVA), with time of day during prolonged wakefulness as a within-subjects factor and *PER3* polymorphisms as between-subjects factor. Three 2 \times 12 mixed-model repeated-measures ANOVAs were conducted for the personality scores, having time of day as a within-subjects factor

and, respectively, High and Low Psychoticism, Extraversion, and Neuroticism as between-group factors. Normality and heterogeneity of distributions were assessed before entering the data in the ANOVA. Bonferroni *post hoc* analysis was applied when appropriate. Statistical significance was set at .05.

TABLE 1
EYSENCK PERSONALITY SCALE SCORES FOR GROUPS SCORING HIGH AND LOW ON EYSENCK PERSONALITY SCALES AND TOTAL SAMPLE

EPS Scale	High			Low			Total		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>
Psychoticism	19	7.9	1.4	20	2.2	0.6	1-12	4.9	2.4
Extraversion	19	17.7	0.7	14	9.8	1.1	7-19	14.4	2.9
Neuroticism	10	15.3	2.8	10	2.0	1.2	0-20	7.6	4.2

RESULTS

Scores on the Eysenck Personality Scale for High and Low Psychoticism, Extraversion, and Neuroticism groups are shown in Table 1. As expected, statistically significant effects of duration of wakefulness during the experimental session on increased subjective sleepiness (KSS), decreased activation, decreased vigilance, and lower mood (Table 2) were found for all personality groups. Statistically significant effects of group were found only for Extraversion scores. During prolonged wakefulness, the Introverted group had higher KSS scores, greater decreases in vigilance and activation and a lower mean mood rating compared to the Extraverted group (Table 2). Statistically significant interactions of duration of wakefulness and group were found only in Extraversion for vigilance and mood. The Introverted group showed greater decreases in subjective measures of vigilance and mood than did the Extraverted group (Table 2 and Fig. 1).

Characteristics of the three genotype groups (PER3 4\4, PER3 5\5, PER3 4\5) are shown in Table 3. No differences on personality scores or circadian preferences were found among the three genotype groups. Statistically significantly increased subjective sleepiness, decreased vigilance, decreased activation, and lower mood across duration of wakefulness was found in all three genotypes (Table 4). No group effect nor time by group interaction were found.

DISCUSSION

The analysis of personality traits showed that during the period of prolonged wakefulness, the Extraverted group had a significantly lower

TABLE 2
RESULTS OF ANALYSES OF VARIANCE FOR EYSENCK PERSONALITY SCALE SCORES

Scale	Time			Group (High, Low)			Time * Group		
	<i>F</i>	<i>df</i>	<i>ES</i>	<i>F</i>	<i>df</i>	<i>ES</i>	<i>F</i>	<i>df</i>	<i>ES</i>
Karolinska Sleepiness Scale									
Psychoticism	26.9‡	11,36	0.43	2.4	1,36		1.3	11,36	
Extraversion	21.9‡	11,30	0.42	8.5*	1,30	0.22	1.1	11,30	
Neuroticism	12.3‡	11,18	0.41	3.2	1,18		1.4	11,18	
Vigilance									
Psychoticism	21.4‡	11,36	0.37	0.12	1,36		0.8	11,36	
Extraversion	18.5‡	11,30	0.38	16.5‡	1,30	0.35	1.1*	11,30	0.11
Neuroticism	9.2‡	11,18	0.34	3.2	1,18		1.6	11,18	
Activation									
Psychoticism	14.1‡	11,36	0.28	3.7	1,36		0.1	11,36	
Extraversion	13.6‡	11,30	0.31	4.9*	1,30	0.14	1.2	11,30	
Neuroticism	6.5‡	11,18	0.27	0.7	1,18		0.8	11,18	
Mood									
Psychoticism	12.6‡	11,36	0.25	0.3	1,36		1.2	11,36	
Extraversion	13.9‡	11,30	0.32	15.5‡	1,30	0.34	2.7‡	11,30	0.10
Neuroticism	5.2‡	11,18	0.22	2.9	1,18		1.1	11,18	

* $p < .05$, † $p < .01$, ‡ $p < .001$.

mean sleepiness score than the Introverted group. According to Eysenck (1967), introverts and extraverts are characterized by different levels of basal arousal; introverts were theorized to have a higher than “optimal” level of basal arousal and extraverts to have a lower than “optimal” level (Taylor & McFatter, 2003). Several studies have found that extraverts perform worse than introverts on vigilance tasks after sleep deprivation

TABLE 3
CHARACTERISTICS OF PERIOD3 (PER3) GENOTYPES ($N = 70$)

Variable	PER3						$F_{2,66}$	p
	4/4		4/5		5/5			
Sample Distribution	$n = 21, 30\%$		$n = 41, 58.5\%$		$n = 8, 11.5\%$			
	M	SD	M	SD	M	SD		
Age, yr.	24.7	4.8	23.7	2.3	24.6	1.9	0.7	0.5
Psychoticism	5.0	2.0	5.0	2.4	5.0	2.2	0.001	0.9
Extraversion	13.8	3.4	14.8	3.4	14.0	4.7	0.8	0.4
Neuroticism	8.9	3.7	7.3	4.2	5.6	4.4	1.9	0.1
r-MEQ	17.0	5.5	18.0	5.4	18.0	5.4	0.3	0.7

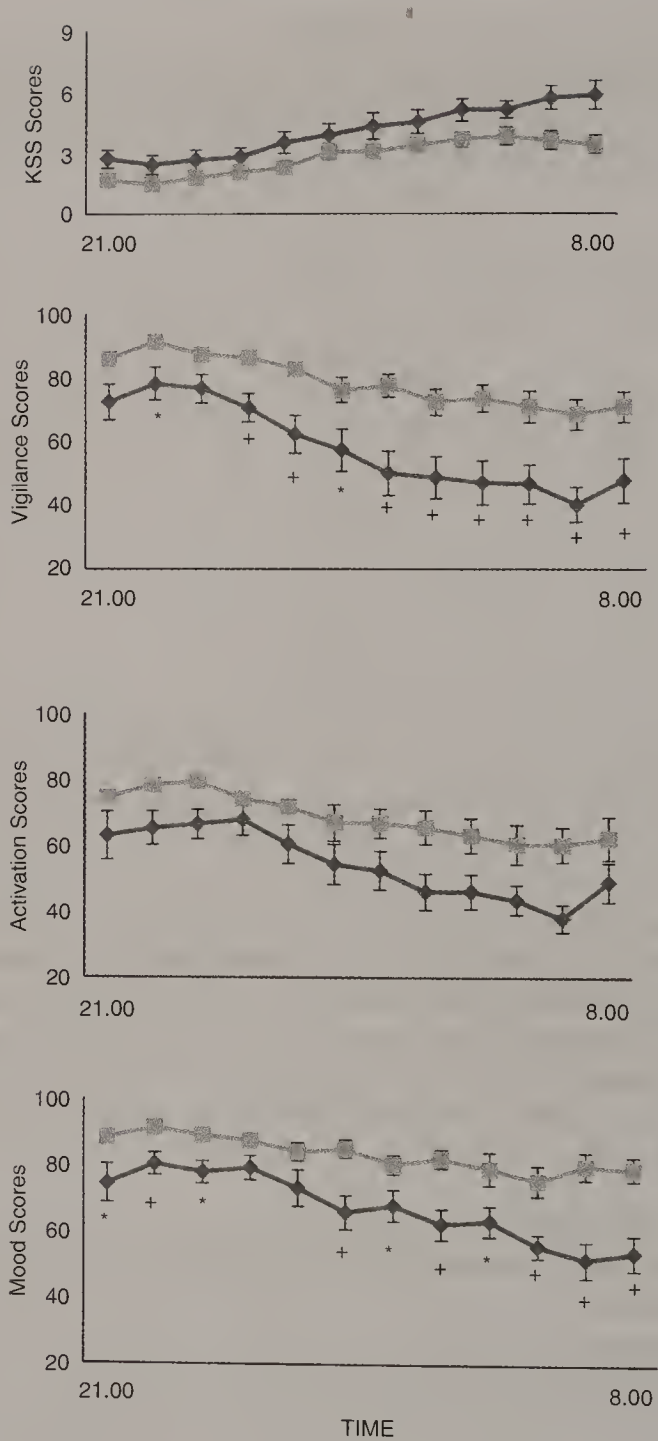


FIG. 1. Visual Analogue Scale (VAS) scores (M , SE) for sleepiness (KSS), vigilance, activation, and mood during prolonged wakefulness in Extraverted (gray lines) and Introverted (black lines) groups. * $p < .05$, † $p < .005$ indicate *post hoc* significant differences between Extraverted and Introverted groups.

TABLE 4
RESULTS OF ANALYSES OF VARIANCE FOR PERIOD3 GENOTYPES

Genotype	Time			Group		Time * Group	
	<i>F</i>	<i>df</i>	<i>ES</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>
Karolinska Sleepiness Scale							
PER3	33.10‡	11,66	0.33	0.51	2,66	0.07	22,66
Vigilance							
PER3	23.26‡	11,66	0.26	1.13	2,66	0.09	22,66
Activation							
PER3	19.49‡	11,66	0.23	0.4	2,66	0.7	22,66
Mood							
PER3	16.02‡	11,66	0.20	0.017	2,66	0.87	22,66

Note.—ES = effect size, eta squared. ‡ $p < .001$.

(Bakan, 1959; Hogan, 1966; Harkins & Geen, 1975; Koelega, 1992; Taylor & McFatter, 2003). Killgore, *et al.* (2007) suggested that extraversive personality traits confer some vulnerability to the adverse neurobehavioral effects of sleep deprivation, whereas introversive traits appear to confer some resistance to these effects. Contrary to Killgore, *et al.* (2007), the present study found that the Introverted group reported more sleepiness during the session. As sleep deprivation continued, the Introverted group's scores on alertness, activity, and mood worsened in a more pronounced manner than in the Extraverted group.

Studies on sleep deprivation have generally been conducted on the assumption that sleep deprivation is a de-arousing condition (Naitoh, Kales, Kollar, Smith, & Jacobson, 1969; Naitoh, Pasnau, & Kollar, 1971; Matthews, 1992); however, a moderate amount of sleep deprivation, as used in the present study, can have instead an arousing effect. As shown by Barbato, Ficca, Beatrice, Casiello, Muscettola, and Rinaldi (1995), during sleep deprivation alpha EEG power decreases while beta EEG power, a measure of arousal, increases. Dumont, Macchi, Carrier, Lafrance, and Hébert (1999) reported that the waking EEG between 18.00-24.75 Hz correlates with subjective sleepiness and may reflect the increasing effort made to stay alert. Total or partial sleep deprivation may also have profound and rapid antidepressant effects in patients with affective disorders (Wirz-Justice & Van den Hoofdakker, 1999). Positive effects of sleep deprivation can be attributed to increase of dopamine cortical arousal, as suggested by recent studies (Barbato, Ficca, Muscettola, Fichelle, Beatrice, & Rinaldi, 2000; Barbato, De Padova, Paolillo, Arpaia, Russo, & Ficca, 2007; Volkow, Wang, Telang, Fowler, Logan, Wong, *et al.*, 2008; Volkow, Tomasi, Wang, Telang, Fowler, Wang, *et al.*, 2009). In addition, Barbato, *et al.* (1995, 2007)

have reported that eye blink rate increases after prolonged wakefulness, suggesting an activation of the dopamine system to maintain wakefulness and counteract a rising sleep drive. Volkow, *et al.* (2008) have shown a significant reduction of [^{11}C]Raclopride in the striatum and thalamus after sleep deprivation, and that the magnitude of this reduction correlated with increases in fatigue and with deterioration in cognitive performance. Thus, dopamine activity increases may occur to maintain arousal as the drive to sleep increases. A dopamine-related increase of arousal can thus be hypothesized following moderate amounts of sleep deprivation.

Consistent with Eysenck's bimodal distribution of extraversion-introversion, the Extraverted group, theorized to have a lower basal arousal, were less affected by the experimental condition. It should also be considered that during the night of sleep deprivation, extraverts may have engaged more than introverts in socialization with other volunteers, thus possibly benefiting of the social component of the experimental condition (Rupp, Killgore, & Balkin, 2010).

A limitation of the study was that only subjective measures were assessed. It is not possible to say whether personality traits might play a role in performance. However, studies that have examined the relationship between subjective and objective sleepiness have found low or non-significant correlations between subjective and objective measures (Benbadis, Mascha, Perry, Wolgamuth, Smolley, & Dinner, 1999). Moreover, as suggested by Johnson, Freeman, Spinweber, and Gomez (1991), subjective and objective measures may reflect different aspects of sleepiness. Large differences between subjective and objective measures in response to sleep deprivation have also been reported. Franzen, Siegle, and Buysse (2008) investigated the effects of sleep deprivation on subjective and objective measures of sleepiness, mood, and psychomotor vigilance performance (PVT). They found that psychomotor vigilance was independent of the sleepiness and affect outcomes. According to their results, objective and subjective measures represent distinct entities that should not be assumed to be equivalent.

Consistent with previous studies (Dijk & Archer, 2010), the PER3 5\5 genotype was found in 11.4% of participants, whereas the PER3 4\5 and PER3 4\4 genotypes were found in 58.6% and 11.30% of participants, respectively. In agreement with Goel, *et al.* (2009), no differences in personality traits or circadian preferences were found among the three genotypes PER3 VNTR (PER3 4\4, PER3 4\5 and PER3 5\5).

Groups with all three genotypes showed increased subjective sleepiness during prolonged wakefulness. Lower mood and a significant decrease in vigilance were also reported by all three genotypes. The results do not support a higher subjective behavioural vulnerability of the five-repeat allele PER3 to prolonged wakefulness. Previous studies of Viola, *et al.*

(2007) have reported that the 5-repeat allele (PER3 5\5) is associated with worse cognitive performance and with higher sleep propensity before and after total sleep deprivation. However, Goel, *et al.* (2009) found that people with the three genotypes show robust and equivalent cumulative decreases in cognitive performance and physiological alertness, as well as cumulative increases in sleepiness, during chronic partial sleep deprivation. Differences in the type of task can also influence the individual responses to sleep deprivation. Groeger, Viola, Lo, von Schantz, Archer, and Dijk (2008) found that people with the longer repeat allele variant of PER3 perform worse than those who are homozygous for the shorter variant, but only on executive tasks, and this was also mediated by homeostatic sleep pressure.

The limited number of individuals in the PER3 5\5 group, the larger representation of women, and the young age of the sample could have masked a biological sensitivity to the experimental sleep deprivation setting. The negative finding should be considered cautiously and further studies are needed to assess the role of genetic factors in mediating individual responses to sleep-wake manipulation. The present results did support Eysenck's original hypothesis that personality traits and their underlying psychobiological systems react differently to manipulation of the arousal system. The prolonged wakefulness used in the present study is a useful method for investigating such psychobiological systems.

REFERENCES

- ADAN, A., & ALMIRALL, H. (1992) The influence of age, work schedule and personality on morningness dimension. *International Journal of Psychophysiology*, 12(2), 95-99.
- ÅKERSTEDT, T., & GILLBERG, M. (1990) Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29-37.
- BAKAN, P. (1959) Extraversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology*, 50(4), 325-332.
- BARBATO, G., DE PADOVA, V., PAOLILLO, A. R., ARPAIA, L., RUSSO, E., & FICCA, G. (2007) Increased spontaneous eye blink rate following prolonged wakefulness. *Physiology & Behavior*, 90(1), 151-154.
- BARBATO, G., FICCA, G., BEATRICE, M., CASIELLO, M., MUSCETTOLA, G., & RINALDI, F. (1995) Effects of sleep deprivation on spontaneous eye blink rate and alpha EEG power. *Biological Psychiatry*, 38, 340-341.
- BARBATO, G., FICCA, G., MUSCETTOLA, G., FICHELE, M., BEATRICE, M., & RINALDI, F. (2000) Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research*, 93(2), 145-151.
- BENBADIS, S. R., MASCHA, E., PERRY, M. C., WOLGAMUTH, B. R., SMOLLEY, L. A., & DINNER, D. S. (1999) Association between the Epworth Sleepiness Scale and the multiple sleep latency test in a clinical population. *Annals of Internal Medicine*, 130(4, Part 1), 289-292.
- CALDWELL, J. A., MU, Q., SMITH, J. K., MISHORY, A., CALDWELL, J. L., PETERS, G., BROWN, D. L., & GEORGE, M. S. (2005) Are individual differences in fatigue vulnerability related to baseline differences in cortical activation? *Behavioral Neuroscience*, 119(3), 694.

- CORCORAN, D. W. J. (1962) Individual differences in performance after loss of sleep. Unpublished doctoral dissertation, Univer. of Cambridge.
- CORCORAN, D. W. J. (1972) Studies of individual differences at the applied psychology unit. In V. D. Nebylitsin & J. A. Gray (Eds.), *Biological bases of individual behavior*. New York: Academic Press. Pp. 269-290.
- DIJK, D. J., & ARCHER, S. N. (2010) PERIOD3, circadian phenotypes, and sleep homeostasis. *Sleep Medicine Review*, 14(3), 151-160.
- DINGES, D., & KRIBBS, N. (1991) Performing while sleepy: effects of experimentally induced sleepiness. In T. Monk (Ed.), *Sleep, Sleepiness and Performance*. Chichester, UK: Wiley. Pp. 97-128.
- DUMONT, M., MACCHI, M. M., CARRIER, J., LAFRANCE, C., & HÉBERT, M. (1999) Time course of narrow frequency bands in the waking EEG during sleep deprivation. *Neuroreport*, 10(2), 403-407.
- DURMER, J. S., & DINGES, D. F. (2005) Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, 25, 117-129.
- EBISAWA, T., UCHIYAMA, M., KAJIMURA, N., MISHIMA, K., KAMEI, Y., KATOH, M., WATANABE, T., SEKIMOTO, M., SHIBUI, K., KIM, K., KUDO, Y., OZEKI, Y., SUGISHITA, M., TOYOSHIMA, R., INOUE, Y., YAMADA, N., NAGASE, T., OZAKI, N., OHARA, O., ISHIDA, N., OKAWA, M., TAKAHASHI, K., & YAMAUCHI, T. (2001) Association of structural polymorphisms in the human PERIOD3 gene with delayed sleep phase syndrome. *EMBO Reports*, 2(4), 342-346.
- EYSENCK, H. J., & EYSENCK, S. (1964) *Manual of the EPI*. Univer. of London.
- EYSENCK, H. J. (1967) *The biological basis of personality*. Springfield, IL: C. C. Thomas.
- FRANZEN, P. L., SIEGLE, G. J., & BUYSE, D. J. (2008) Relationships between affect, vigilance, and sleepiness following sleep deprivation. *Journal of Sleep Research*, 17(1), 34-41.
- GOEL, N., BANKS, S., MIGNOT, E., & DINGES, D. F. (2009) PER3 polymorphism predicts cumulative sleep homeostatic but not neurobehavioral changes to chronic partial sleep deprivation. *Public Library of Science*, 4(6), e5874.
- GROEGER, J. A., VIOLA, A. U., LO, J. C., VON SCHANTZ, M., ARCHER, S. N., & DIJK, D. J. (2008) Early morning executive functioning during sleep deprivation is compromised by a PERIOD3 polymorphism. *Sleep*, 31(8), 1159-1167.
- HARKINS, S., & GEEN, R. G. (1975) Discriminability and criterion differences between extraverts and introverts during vigilance. *Journal of Research in Personality*, 9(4), 335-340.
- HARRISON, Y., & HORNE, J. (1998) Sleep loss impairs short and novel language tasks having prefrontal focus. *Journal of Sleep Research*, 7(2), 95-100.
- HEISER, P., DICKHAUS, B., OPPER, C., HEMMETER, U., REMSCHMIDT, H., WESEMANN, W., KRIEG, J. C., & SCHREIBER, W. (2001) Alterations of host defense system after sleep deprivation are followed by impaired mood and psychosocial functioning. *World Journal of Biological Psychiatry*, 2(2), 89-94.
- HOGAN, M. J. (1966) Influence of motivation on reactive inhibition in extraversion-introversion. *Perceptual & Motor Skills*, 22, 187-192.
- JOHNSON, L. C., FREEMAN, C. R., SPINWEBER, C. L., & GOMEZ, S. A. (1991) Subjective and objective measures of sleepiness: effect of benzodiazepine and caffeine on their relationship. *Psychophysiology*, 28(1), 65-71.
- JOHNSON, L. C., & MACLEOD, W. L. (1973) Sleep and awake behaviour during gradual sleep reduction. *Perceptual & Motor Skills*, 36, 87-97.
- KAIDA, K., TAKAHASHI, M., ÅKERSTEDT, T., NAKATA, A., OTSUKA, Y., HARATANI, T., & FUKASAWA, K. (2006) Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7), 1574-1581.

- KILLGORE, W. D., RICHARDS, J. M., KILLGORE, D. B., KAMIMORI, G. H., & BALKIN, T. J. (2007) The trait of introversion-extraversion predicts vulnerability to sleep deprivation. *Journal of Sleep Research*, 16(4), 354-363.
- KOELEGA, H. S. (1992) Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological Bulletin*, 112(2), 239-258.
- MATTHEWS, G. (1992) Extraversion. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance: state and trait*. Vol. 3. London, UK: Academic Press.
- MONK, T. H., & EMBREY, D. E. (1989) A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research*, 27, 89-99.
- MONK, T. H., REYNOLDS, C. F., III, KUPFER, D. J., BUYSSE, D. J., COBLE, P. A., HAYES, A. J., MACHEN, M. A., PETRIE, S. R., & RITENOUR, A. M. (1994) The Pittsburgh Sleep Diary. *Journal of Sleep Research*, 3, 111-120.
- NAITOH, P., KALES, A., KOLLAR, E. J., SMITH, J. C., & JACOBSON, A. (1969) Electroencephalographic activity after prolonged sleep loss. *Electroencephalography and Clinical Neurophysiology*, 27, 2-11.
- NAITOH, P., PASNAU, R. O., & KOLLAR, E. J. (1971) Psychophysiological changes after prolonged deprivation of sleep. *Biological Psychiatry*, 3, 309-320.
- NATALE, V. (1999) Validazione di una Scala ridotta di mattutinità (r-MEQ) [Validation of a reduced scale of Morningness-Eveningness]. *Bollettino di Psicologia Applicata*, 229, 19-26. [in Italian]
- RUPP, T. L., KILLGORE, W. D. S., & BALKIN, T. J. (2010) Socializing by day may affect performance by night: vulnerability to sleep deprivation is differentially mediated by social exposure in extraverts vs introverts. *Sleep*, 33(11), 1475-1485.
- TAYLOR, D. J., & MCFATTER, R. M. (2003) Cognitive performance after sleep deprivation: does personality make a difference? *Personality and Individual Differences*, 33, 1179-1193.
- VAN DONGEN, H. P. A., BAYNARD, M. D., MAISLIN, G., & DINGES, D. F. (2004) Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, 27(3), 423-433.
- VAN DONGEN, H. P. A., VITELLARO, K. M., & DINGES, D. F. (2005) Individual differences in adult human sleep and wakefulness: leitmotif for a research agenda. *Sleep*, 28(4), 479-496.
- VIOLA, A. U., ARCHER, S. N., JAMES, L. M., GROEGER, J. A., LO, J. C., SKENE, D. J., VON SCHANTZ, M., & DIJK, D. J. (2007) PER3 polymorphism predicts sleep structure and waking performance. *Current Biology*, 17(7), 613-618.
- VOLKOW, N. D., TOMASI, D., WANG, G-J., TELANG, F., FOWLER, J. S., WANG, R. L., LOGAN, J., WONG, C., JAYNE, M., & SWANSON, J. M. (2009) Hyperstimulation of striatal D2 receptors with sleep deprivation: implications for cognitive impairment. *NeuroImage*, 45(4), 1232-1240.
- VOLKOW, N. D., WANG, G-J., TELANG, F., FOWLER, J. S., LOGAN, J., WONG, C., MA, J., PRADHAN, K., TOMASI, D., THANOS, P. K., FERRÉ, S., & JAYNE, M. (2008) Sleep deprivation decreases binding of [¹¹C]raclopride to dopamine D2/D3 receptors in the human brain. *Journal of Neuroscience*, 28(34), 8454-8461.
- WIRZ-JUSTICE, A., & VAN DEN HOOFDACKER, R. H. (1999) Sleep deprivation in depression: what do we know, where do we go? *Biological Psychiatry*, 46(4), 445-453.

UP- AND DOWN-REGULATION OF DAILY EMOTION: AN EXPERIENCE SAMPLING STUDY OF CHINESE ADOLESCENTS' REGULATORY TENDENCY AND EFFECTS^{1, 2}

XINMEI DENG, BIAO SANG, AND ZIYAN LUAN

*School of Psychology and Cognitive Science
East China Normal University, Shanghai*

Summary.—The present study examined Chinese adolescents' emotion regulatory tendency and its effect, using an Experience Sampling Method. Participants comprised 72 Chinese adolescents (M age = 15.2 yr., SD = 1.7; 36 girls). Momentary emotional experience and regulation was assessed up to 5 or 6 times each day for two weeks. Results showed that participants tended to use up-regulation when they experienced positive emotion and habitually regulated their negative emotion by down-regulation. Also, adolescents who utilized down-regulation in a certain sampling moment reported higher positive emotion at the subsequent sampling moment. Moreover, adolescents who utilized down-regulation more frequently reported higher positive emotion at the subsequent sampling moment. Overall, down-regulation seemed to be a more adaptive regulatory strategy than up-regulation in Chinese adolescents' emotional lives.

Adolescents' emotional life has been depicted as a period full of "storm and stress" (Casey, Jones, Levita, Libby, Pattwell, Ruberry, *et al.*, 2010). Adolescents' daily emotional experiences and use of relevant emotion-regulation strategies are remarkably varied among individuals and depend on the demands of context (Coats & Blanchard-Fields, 2008; Matsumoto, Yoo, Nakagawa, & 37 Members Multinational Study of Cultural Display Rules, 2008). The most widely accepted definition of emotion regulation was proposed by Gross and Levenson (1993), who suggested that the process of emotion regulation is an integration of physiological, cognitive, and behavior processes. Emotion-regulation strategies comprise up-regulation and down-regulation. Up-regulation is defined as strategies to increase and maximize the emotional experience and amplify behavioral and facial responses. Conversely, down-regulation is defined as strategies to decrease and minimize the intensity of emotional experience, and weaken behavioral and facial responses (Krompinger, Moser, & Simons, 2008). Different strategies give rise to different regulatory consequences.

¹Address correspondence to Biao Sang, East China Normal University, Shanghai, 200062, P. R. China or e-mail (bsang@psy.ecnu.edu.cn).

²This work was financially supported by the National Natural Science Foundation [No.31371043], the Ministry of Education of Humanities, the Social Science Research Base on Major Projects of China [10JJDXXL001], and the Ministry Reward for Excellent Doctors in Academics.

Previous studies have indicated that habitual usage of down-regulation reduces negative effects to the cardiovascular system of experiencing intensely negative emotion (Giuliani, McRae, & Gross, 2008), and improves one's ability to regulate emotion (Gross, 2001). Usage of up-regulation may increase blood pressure (Butler, Lee, & Gross, 2009) and prolong the experience of negative emotions (Rusting & Nolen-Hoeksema, 1998).

Although previous studies have examined the consequences of different emotional regulatory strategies to daily emotional experience, some issues still remain unclear. Firstly, most studies have focused on the regulation of negative emotion (Sivers, McRae, Gabrieli, Gross, Remy, & Ochsner, 2012) without assessing the regulation of positive emotion. However, the regulation of positive emotion is indispensable to capturing the whole picture of emotion regulation, and this is especially important in certain cultural contexts. For example, in the Chinese collectivistic culture, which is heavily influenced by Confucianism, the goal of maintaining interpersonal harmony gives rise to avoiding the expression of extreme happiness in front of others. Down-regulating positive emotions protects adolescents from being criticized as arrogant or inconsiderate of others' feelings (Miyamoto & Ma, 2011).

Emotion regulation is also a dynamic and time-related process (Cole, Martin, & Dennis, 2004). The contextual and responsive nature of emotion implies that emotional experience changes dynamically (Kuppens, Oravecz, & Tuerlinckx, 2010). Emotion regulation must be contextualized, since situational factors may influence the regulation goal at specific moments (Novin, Rieffe, & Mo, 2010). Reliable and contingent patterns of emotion regulation, in particular, can be better understood if researchers examine emotion regulation in procedures eliciting real social interactions (Cole, *et al.*, 2004). The Experience Sampling Method (ESM) is effective in studying emotion regulation and its effects on daily emotional experience. Researchers collect data from participants at selected moments during their normal daily activities (Csikszentmihalyi & Larson, 1987; Larson, 1989; Silk, Steinberg, & Morris, 2003). Participants receive a signal (e.g., a "beep" from a device) and then report their thoughts, events, experiences, and regulation strategies on the Experience Sampling Form (ESF). The first advantage is that such measurements reduce recall bias (Scollon, Diener, Oishi, & Biswas-Diener, 2004) and decrease the effects of implicit attitudes (Conner & Barrett, 2005); e.g., when asked to report their habitual regulation strategy, people might report the one they "think" they would use or ought to use (Mauss, Evers, Wilhelm, & Gross, 2006). The second advantage of ESM is that it can capture the habitual usage of emotion regulation in a naturalistic setting (Csikszentmihalyi & Larson, 1987; Larson, 1989). Artificial emotion-eliciting situations might not activate emotion

and emotion regulation in the same way and to the same extent as real-life situations (Schimmack, 2001, 2003). ESM assesses daily emotion regulation in a naturalistic setting where people use their habitual styles of emotion regulation.

When exploring the effects of different emotion regulatory strategies, it is important to consider not only the fulfillment of short-term hedonic benefits, but also to consider the fulfillment of the individual's longer-term benefits from the strategy (Koole, 2009). Prior studies have suggested that individuals want to experience emotions not only for their short-term hedonic benefit, but also for their longer-term instrumental benefits (Tamir, 2005; Tamir, Chiu, & Gross, 2007; Tamir, Mitchell, & Gross, 2008). Although down-regulating negative emotions (e.g., re-appraising sadness and distracting one's attention from negative stimuli) are suggested, intentionally increasing certain negative emotions also seems to be encouraged for long-term benefits (Tamir, *et al.*, 2007). Similarly, although using the up-regulation strategy in positive emotional conditions increases short-term positivity of emotional experience, down-regulating positive emotion might avoid potential negative outcomes in the future (Soto, Perez, Kim, Lee, & Minnick, 2011).

The Present Study

As mentioned above, the Experience Sampling Method is an alternative to the ordinary memory-based measures, survey-based measures, and laboratory measures for momentary assessment. It is less susceptible to retrospective bias and has higher ecological validity. The present study adopted this methodology to explore adolescents' daily emotion-regulation strategies towards positive as well as negative emotions, and their effects. The first goal was to assess the tendency of regulatory use of these Chinese adolescents (Hypothesis 1). The second goal was to identify the effects of different regulatory use in adolescents' daily life (Hypotheses 2 & 3).

Hypothesis 1. Generally, adolescents were expected to have a hedonic tendency to maximize positive emotion and minimize negative emotion (Miyamoto & Ma, 2011; Webb, Miles, & Sheeran, 2012). That is to say, when confronting positive emotional episodes, adolescents would be more likely to use up-regulation, whereas in response to negative emotional episodes, adolescents would be more likely to use down-regulation.

Hypothesis 2. Given the different consequences of up- and down-regulation mentioned above, adolescents who reported up-regulation less frequently were expected to report a higher rating of the positivity of their emotional experience. Adolescents who reported more frequent down-regulation were

also expected to report a higher rating of the positivity of their emotional experience.

Hypothesis 3. Considering the long-term consequences of up-regulation and down-regulation, more than simply hedonic emotion regulation was expected. Specifically, adolescents who reported down-regulation at a specific sampling time were expected to report a higher rating of positivity of their emotional experience at the subsequent sampling time. Adolescents who utilized up-regulation at a certain sampling time were expected to report a lower rating of positivity of their emotional experience at the subsequent sampling time.

METHOD

Participants

Data utilized for this analysis were collected at 11 public schools in Shanghai. Before the school assembly, researchers contacted all of the partner schools³ in Shanghai and provided them with information about the study, inviting participation. Eleven schools permitted recruitment of participants, who then were randomly selected. Participants comprised 72 adolescents (36 girls). Ages ranged between 12 and 18 years ($M = 15.2$, $SD = 1.7$) and all were Chinese (87.5% from single-child homes). Approximately 9.7% of the fathers and 15.3% of the mothers had not completed a high school education; 33.3% of the fathers and 26.4% of the mothers had received a college education.

Measures

Emotional experience.—Assessment of emotional experience was adopted from the work of Larson, Moneta, Richards, and Wilson (2002). Students were asked to indicate their emotional experiences (current, peak, and general emotional experiences) in a specially designed booklet every 90 minutes during the school day. The students rated their current, peak, and general emotional experiences on a 7-point Likert-type scale with anchors -3 : Very negative emotion and $+3$: Very positive emotion. Current emotional experience was elicited by asking, "How are you feeling now?" at each sampling moment. Peak emotional experience was defined as the most intense emotional experience since the last sampling moment, either positive or negative. Participants were asked, "What was your most intense feeling in the past one and a half hours?" followed by an open-ended

³A partner school is a school that cooperates with the authors' department for research studies. In return, the department offers teacher-training classes to these partner schools. Partner schools also provide internship opportunities to students in the department. There are 11 partner schools. All of the 11 public schools in the present study are partner schools.

question, "What happened?"⁴ General emotional experience was an evaluation of the emotional experience of the participants since the last sampling moment; participants were asked "Generally speaking, how did you feel in the past one and a half hours?" Cronbach's α for the rating of three Likert-type scale ratings of emotional experiences was .78 and the split-half correlation coefficient was .70.

Emotion regulation.—At each sampling moment, participants also assessed the corresponding emotion-regulation strategies they used in response to their current and peak emotional experiences. General emotional experiences were also assessed. To collect more information, open-ended questions about emotion-regulation strategies were used instead of a strategy checklist. For example, when asking the question corresponding to peak emotional experience, participants were asked, "What was your most intense feeling in the past one and a half hours? Did you do anything about it at the time? If so, what?" Two independent raters who were blind to the hypotheses categorized narratives into three categories of regulatory strategy: up-regulation, down-regulation, and no regulation (Giuliani, *et al.*, 2008; Krompinger, *et al.*, 2008). Inter-rater reliability was calculated for the consistency of categorization (Cohen's $\kappa = .88$).

Demographic information.—Information assessed included participants' age, gender, student ID, grade, paternal and maternal education levels, and family structure (e.g., "Are you a single child? If not, how many siblings do you have? With whom are you living?"). No demographic index had any statistically significant association with the variables of interest ($p > .05$).

Procedure

During assemblies in each school, participants were informed about the sampling procedures and that they would be participating in the study for two weeks. Research assistants provided each participant with a booklet containing a description of the study, a parental consent form, and a self-consent form. All participants returned the parental consent and the self-consent forms.

After the study began, each participant was asked to report their emotional experiences (current, peak, and general) and corresponding emotion-regulation strategies five or six times each day according to the school schedule, within every 90-min. block during school hours (7:30 to 16:30). Data collection occurred only on weekdays, for two weeks. According to this design, each participant was to report 55 times. In the present study, data from 3,335 sampling moments were collected, for a compliance rate

⁴Emotional experience and its corresponding emotional events were examined with the open-ended question ("What happened?"); participants replied on a few blank lines.

of 84.2%. The minimum number of samples was 27 (49.09% of the planned samples).

Analysis

Cross-tabulations of emotion-regulation strategies and emotional experience ratings were done to describe the tendency of regulatory strategy use (up- or down-regulation) of Chinese adolescents when they reported positive and negative emotional episodes (Hypothesis 1). Hierarchical linear modeling (HLM) was conducted to investigate the strength of associations of using each type of emotion-regulation strategy on the rating of the emotional experience 90 min. later (Hypotheses 2 and 3). HLM is widely used to analyze data comprising repeated-measures sampling. It has the advantage of providing information about the tendency of regulatory use in daily emotional lives on the intra-individual level (Klumb, Elfering, & Herre, 2009). The first HLM estimated the contribution of emotion-regulation strategy (up-regulation and down-regulation) to emotional experience ratings. Another two models were run to estimate the lagged relationships between emotion-regulation strategy (at T_{n-1}) and later emotional experience ratings (at T_n). SPSS 17.0 and HLM 6.08 were used for data analysis.

RESULTS

Regulatory Tendency of Daily Emotional Experience of Chinese Adolescents

The sampling moments of interest were those in which participants reported at least a moderate emotional intensity rating (-3 or -2) for negative emotion or for positive emotion (2 or 3). Like many laboratory experiments, the present experience sampling study aimed to examine adolescents' emotional responses in a condition of at least moderate emotional intensity. To keep participants in a relatively emotionally aroused condition, participants were exposed to moderately or highly emotionally aroused pictures or videos. It is easier to observe the tendency and effect of specific regulatory strategies in a moderately or highly emotionally aroused condition. Likewise, the present study analyzed the tendency and effect of regulatory use in those moments when participants reported at least a moderate emotional intensity rating. About one-quarter of the data met these criteria (852 of 3,335) and were included in the analyses.⁵

Cross-tabulations of emotion-regulation strategies and emotional experience ratings are presented in Table 1. Although Chinese adolescents reported more down-regulatory strategy use (66.4%) than up-regulatory strategy use (33.6%), results differed for positively and negatively rated emotional epi-

⁵Here we wanted to analyze the effects of different regulatory use the changes of emotional experience. Therefore, values of the changes of emotional experience are more important than the values of emotional experience in a single sampling moment.

TABLE 1
CROSS-TABULATION OF EMOTION-REGULATION STRATEGY AND EMOTIONAL EXPERIENCE RATINGS

Emotion Regulation	Positive Emotion		Negative Emotion		Total	
	Observed	Expected	Observed	Expected	<i>n</i>	%
Down-regulation	90	105	167	152	257	66.4
Up-regulation	66	51	59	74	125	33.6

Note.—Emotion Regulation × Emotional Experience: LR (Likelihood Ratio) = 10.92; *df* = 4, *p* < .001.

sodes. When reporting positively rated emotional episodes, Chinese adolescents were more likely to report using up-regulation, whereas in response to negatively rated emotional episodes, the adolescents were more likely to report using down-regulation, supporting Hypothesis 1.

Mathematical Modeling of the Effect of Emotion Regulation on Emotional Experience

Hierarchical linear models (HLM) were used to account for the possible influence of emotion-regulation strategy on later emotional experience ratings. A two-level data structure was constructed at the sampling level and the personal level:

$$\begin{aligned} \text{Sampling level: } EE_{ij} &= \beta_{0j} + \beta_{1j} \times ER_{ij} + \varepsilon_{ij}, \\ \text{Personal level: } \beta_{0j} &= \gamma_{00} + \mu_{0j}, \beta_{1j} = \gamma_{10} + \mu_{1j}, \end{aligned}$$

where EE and ER represent the emotional experience rating and emotion-regulation strategy, respectively. The regression equation at the sampling level related the EE scores of the *i*th participant at the *j*th sampling moment to their emotion-regulation strategy, where β_{0j} is the grand mean value of EE at times of down-regulation, β_{1j} is the mean effect of emotion-regulation strategy, the coefficient ε_{ij} is the random sampling-level factor that represents the participant's score's deviations from the overall scores, the parameters γ_{00} and γ_{10} are participant-specific random regression coefficients, and μ_{0j} and μ_{1j} are sampling-level residual terms. The analysis was only based on the sampling moments in which emotion-regulation strategies were reported and participants reported at least a moderate emotional intensity rating for negative emotion (−3 or −2) or for positive emotion (2 or 3). Table 2 shows that emotion-regulation strategy was statistically significantly related to the general emotional experience rating ($\gamma_{10} = -.84, p < .05$). Adolescents who less frequently reported up-regulation had higher positive emotional experience ratings. In contrast, adolescents who utilized down-regulation more frequently reported higher positive emotional experience ratings, supporting Hypothesis 2.

TABLE 2
HIERARCHICAL LINEAR MODELING FOR UP-REGULATION AND DOWN-REGULATION STRATEGIES

	Effect	β	95%CI β	SE	df	t
General	γ_{10}	-.61	-1.00, -.022	.20	61	-3.05†
emotional experience	γ_{20}	-.84	-1.51, -.017	.34	61	-2.45*

* $p < .05$, † $p < .01$.

*Lagged Relationships Between Emotion Regulatory Strategies and
Later Emotional Experience*

To investigate the mean effect of emotion-regulation strategies on emotional experience ratings, data were time-lagged and two more HLMs were conducted. The lagged analysis was based on all the sampling moments with emotion regulation-strategies reported. Instead of focusing on the actual values of emotional experience in a single sampling moment, the changes of the rating of emotional experience were used to examine the lagged effects of different regulatory strategy use. Results estimated the lagged relationship between emotion-regulation strategies and emotional experience ratings in the next sampling moment, approximately 90 min. later (Table 3).⁶ To differentiate between up- and down-regulation of positive and negative emotional experiences, analysis was done twice, once with data having negative ratings and then again with data having positive ratings.

Sampling level: $EE_n = \beta_{0j} + \beta_{1j} \times ER_{n-1} + \beta_{2j} \times ER_{n-1} + \varepsilon_{ij}$,
Personal level: $\beta_{0j} = \gamma_{00} + \mu_{0j}, \beta_{1j} = \gamma_{10} + \mu_{1j}, \beta_{2j} = \gamma_{20} + \mu_{2j}$,

where EE_{n-1} represents the emotional experience rating at T_{n-1} , ER_{n-1} represents emotion-regulation strategy use at T_{n-1} , and EE_n represents the emotional experience rating at T_n . The regression equation at the sampling level related the EE scores at T_n to their emotion-regulation strategy and emotional experience rating at the previous assessment (T_{n-1}), where β_{0j} is the grand mean value of EE at times of down-regulation at the first sampling moment, β_{1j} is the mean effect of emotional experience ratings per emotion-regulation strategy in the next sampling moment, β_{2j} is the mean effect of emotional experience to emotional experience in the next sampling moment, the coefficient ε_{ij} is random sampling-level factor representing the participant's score's deviations from the overall scores, the parameters γ_{00} , γ_{10} and γ_{20} are participant-specific random regression coefficients, and μ_{0j} , μ_{1j} and μ_{2j} are sampling-level residual terms.

⁶If the subsequent moment's data were not available, the data pair was omitted from the analysis.

TABLE 3
A TIME-LAGGED HIERARCHICAL LINEAR MODEL FOR UP-REGULATION AND DOWN-REGULATION STRATEGIES IN POSITIVE AND NEGATIVE EMOTIONAL CONDITIONS

Emotional Experience	Effect	β	95%CI β	SE	df	t
Positive emotional experience	γ_{00}	.30	.12, .48	.09	47	3.22†
	γ_{10}	-.48	-.81, -.15	.17	47	-2.86†
	γ_{20}	.20	.02, .38	.09	47	2.11*
Negative emotional experience	γ_{00}	-.16	-.34, -.02	.09	52	-1.69
	γ_{10}	-.21	-.88, .46	.34	52	-0.62
	γ_{20}	.34	.12, .56	.11	52	3.09†

* $p < .05$, † $p < .01$.

In the positive emotional condition, emotional experience ratings at T_{n-1} were positively associated with the emotional experience ratings at the subsequent assessment (T_n ; $\gamma_{20} = .20$, $p < .05$). Emotion-regulation strategy at T_{n-1} was statistically significantly related with the emotional experience ratings at the subsequent assessment (T_n ; $\gamma_{10} = -.48$, $p < .001$). In the positive emotional condition, adolescents who reported down-regulation at T_{n-1} had higher positive emotion ratings at T_n . In contrast, adolescents who reported up-regulation at T_{n-1} had lower positive emotion ratings at T_n . In the negative emotional condition, emotional experience ratings at T_{n-1} were positively associated with the emotional experience ratings at the subsequent assessment (T_n ; $\gamma_{20} = .34$, $p < .01$). However, emotion-regulation strategy at T_{n-1} was not statistically significantly related with the emotional experience ratings at the subsequent assessment (T_n ; $\gamma_{10} = -.21$, $p > .05$).

DISCUSSION

The present study is the first to examine adolescents' emotion-regulation tendencies and their effects on emotion ratings in Chinese adolescents. It is also the first study to adopt the experience sampling method (ESM) for use in a Chinese adolescent population. Results suggested that adolescents tended to use more down-regulatory strategies (66.4%) than up-regulatory strategies (33.6%). In Chinese culture, down-regulation is in keeping with the culturally appropriate goal of minimizing the consequences of emotional experiences. Findings also indicated that Chinese adolescents typically chose to use up-regulation when they experienced positive emotion, and down-regulation when they experienced negative emotion. This suggests that these Chinese adolescents have a hedonic tendency to maximize positive emotion and minimize negative emotion. It is possible that this regulatory tendency is acquired naturally and is influenced by the desire

for friendly interactions and the avoidance of conflict. These results were consistent with Hypothesis 1.

However, up-regulating positive emotion at specific times was not associated with overall greater positive emotion. Hedonic motivation may not always result in an overall adaptive consequence. To support this, effects of different regulatory strategy use were examined in the later HLM analyses. The main finding from the HLM analyses suggested that adolescents who utilized down-regulation more frequently reported higher ratings of positive emotional experience in their daily lives. Adolescents who less frequently reported up-regulation also had higher ratings of positive emotional experience. In addition, the result of lagged analyses between emotional regulatory strategies and later emotional experience provided evidence that the outcome of emotion regulation was related to the regulatory strategies used 90 min. previously. When regulating positive emotional experience, adolescents who reported up-regulation at T_{n-1} had lower positive emotion ratings at T_n . By contrast, adolescents who reported down-regulation at T_{n-1} had higher positive emotion ratings at T_n .

There are two possible ways by which down-regulation may be associated with overall greater subsequent positive emotional experience. Firstly, down-regulation could increase individuals' emotion-regulation skills. Previous studies have suggested that habitual down-regulation is related to increased emotional attractor strength that could help individuals regulate their emotions back to a desired or optimal affective state (Gross, 2001; Kuppens, *et al.*, 2010). Every time individuals used a down-regulation strategy in response to certain emotional events, they reset their level of emotion. The regulatory process trains greater capability in regulating emotions and increases adjustment. Previous studies have shown that although habitual up-regulation would not decrease the emotional attractor strength on a short time scale, it was characterized by decreased attractor strength in the longer term (Kuppens, *et al.*, 2010). With the findings of previous studies, this might be one of the possible explanations why participants who up-regulated their positive emotions at a certain sampling moment did not have overall greater positive emotion ratings. Of course, whether adolescents' habitual regulatory use directly influences their emotional attractor strength and the consequences of regulation needs to be examined in further studies.

Secondly, as mentioned before, down-regulation of positive emotion fits the Chinese cultural attitude toward emotion regulation. Avoiding extreme positive emotion protects adolescents from being criticized as arrogant or inconsiderate of others' feelings. Previous studies have suggested that attitudes toward emotion regulation influence individuals' regulation strategies (Liu, 2007; Mauss, Cook, & Gross, 2007; Liu & Sang, 2009). Those

who believe that emotion should be regulated and controlled (a “positive regulation attitude”) would automatically regulate their emotions when confronting emotional stimuli; however, this was not found for those who believed that emotion should not be regulated and controlled (negative regulation attitude). In Chinese culture, negative emotional attitudes are considered irrational and unnecessary (Mauss, *et al.*, 2007). For example, in many Chinese historical stories and literature, concealing strong emotions and being calm is seen to be a prerequisite for success. The Chinese phrases “泰然自若” and “荣辱不惊,” which mean “appear to be indifferent confronting honor or humiliation,” are typical descriptions of sophisticated and successful adults. Extreme emotions are considered inappropriate or dangerous, especially positive emotions. Chinese believe extreme joy begets sorrow. Undue expression of positive emotion would lead to harm of interpersonal relationships (Deng, 2009). Down-regulating positive emotions is anti-hedonic, but is socially adaptive under Chinese culture. This might be one explanation of why adolescents who reported down-regulating positive emotional experiences had higher positive emotion ratings at the subsequent sampling moment.

However, a lagged relationship between emotion-regulation strategies and emotional experience was only statistically significant for the positive emotional experience data set. Adolescents who reported down-regulating negative emotional experience did not necessarily have higher positive emotion ratings at the subsequent moment. Also, adolescents who reported up-regulating negative emotional experience did not necessarily have lower positive emotion ratings at the subsequent moment. The reason for the discrepancy might be that up-regulating negative emotions does not harm interpersonal harmony, whereas up-regulating positive emotions could. Up-regulating negative emotions might even lead to certain long-term benefits (Tamir, 2005; Tamir, *et al.*, 2007); for example, to avoid making similar mistakes, Chinese parents and teachers encourage adolescents to remember the lessons from past negative experiences (both academic and interpersonal). Undergoing self-imposed hardships is not only seen as cultivating the virtue of patience, but also is seen as a potent method to future success. The complications resulting from negative emotions in Chinese culture makes the effect of negative emotion regulation diverse and complex in different situations. With this, down-regulating negative emotional experience does not necessarily lead to higher positive emotion ratings, and up-regulating negative emotional experience does not necessarily lead to lower positive emotion ratings in the future.

Limitations and Further Research

Although ESM solves the problems of retrospection and ecological validity, it has unique disadvantages. Repeated-measures of emotional

experience and emotion regulation make participants more sensitive to the phenomenon under investigation, and they may report at a higher frequency across time as a result (Klumb, *et al.*, 2009). Estimating the extent of this influence is the direction of further study. Another limitation concerns the reduced number of items or single-item measures in ESM. At every sampling, participants were asked to report their momentary experience and its corresponding regulation by an open-ended question. It is hard to conduct multivariate statistical analysis with the data from each sampling moment. Hence, utilizing multiple methods to gain convergent results is highly recommended (Cole, *et al.*, 2004). Combining various methodologies, such as experience sampling method, neural imaging techniques, laboratory methods, and longitudinal design to achieve convergent results, is a promising direction for future research.

All in all, adopting ESM to examine the emotion-regulation tendencies and their effects in Chinese adolescents provided a more complete picture of emotion regulation. Down-regulating negative emotion and up-regulating positive emotion may have different influences on emotional experience in the short term than in the long term. Results indicate the importance of gaining contextual and anti-hedonic perspectives of judging the adaptive value of emotion regulation. Therefore, when assessing the effects of different emotion regulatory strategies, researchers should consider not only the fulfillment of hedonic need but also the fulfillment of longer-term instrumental goals. Overall, down-regulating emotion has the adaptive potential to maintain interpersonal harmony and mental health, which makes it an effective and adaptive regulatory strategy among Chinese adolescents.

REFERENCES

- BUTLER, E. A., LEE, T. L., & GROSS, J. J. (2009) Does expressing your emotions raise or lower your blood pressure? The answer depends on cultural context. *Journal of Cross-Cultural Psychology*, 40, 510-517.
- CASEY, B. J., JONES, R. M., LEVITA, L., LIBBY, V., PATTWELL, S. S., RUBERRY, E. J., & SOMERVILLE, L. H. (2010) The storm and stress of adolescence: insights from human imaging and mouse genetics. *Developmental Psychobiology*, 52, 225-235.
- COATS, A. H., & BLANCHARD-FIELDS, F. (2008) Emotion regulation in interpersonal problems: the role of cognitive-emotional complexity, emotion regulation goals, and expressivity. *Psychology and Aging*, 23, 39-51.
- COLE, P. M., MARTIN, S. E., & DENNIS, T. A. (2004) Emotion regulation as a scientific construct: methodological challenges and direction for child development research. *Child Development*, 75, 317-333.
- CONNER, T., & BARRETT, L. F. (2005) Implicit self-attitudes predict spontaneous affect in daily life. *Emotion*, 5, 476-488.
- CSIKSZENTMIHALYI, M., & LARSON, R. (1987) Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Disease*, 175, 526-536.

- DENG, X. M. (2009) Development of meta-emotion and emotion regulation: exploring the relations of interrelationship. Paper presented at the Regional Conference of ISSBD, Nanjing, China. June 12–14, 2009.
- GIULIANI, N. R., MCRAE, K., & GROSS, J. J. (2008) The up- and down-regulation of amusement: experimental, behavioral, and autonomic consequences. *Emotion*, 8, 714–719.
- GROSS, J. J. (2001) Emotion regulation in adulthood: timing is everything. *Current Directions in Psychological Science*, 10, 214–219.
- GROSS, J. J., & LEVENSON, R. W. (1993) Emotional suppression: physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, 64, 970–986.
- KLUMB, P., ELFERING, A., & HERRE, C. (2009) Ambulatory assessment in industrial/organizational psychology: fruitful examples and methodological issues. *European Psychologist*, 14, 120–131.
- KOOLE, S. (2009) The psychology of emotion regulation: an integrative review. *Cognition & Emotion*, 23, 4–41.
- KROMPINGER, J. W., MOSER, J. S., & SIMONS, R. F. (2008) Modulations of the electrophysiological response to pleasant stimuli by cognitive reappraisal. *Emotion*, 8, 132–137.
- KUPPENS, P., ORAVECZ, Z., & TUERLINCKX, F. (2010) Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99, 1042–1060.
- LARSON, R. (1989) Beeping children and adolescents: a method for studying time use and daily experience. *Journal of Youth and Adolescence*, 18, 511–530.
- LARSON, R., MONETA, G., RICHARDS, M. H., & WILSON, S. (2002) Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, 73, 1151–1165.
- LIU, J. S. (2007) Mechanisms and development of implicit attitude to emotion regulation. Unpublished doctoral dissertation, China Academic Library and Information System.
- LIU, J. S., & SANG, B. (2009) Effects of the implicit attitude towards emotion regulation on individual emotion regulation. *Journal of Psychological Science*, 32, 571–574.
- MATSUMOTO, D., YOO, S. H., NAKAGAWA, S., & THIRTY-SEVEN MEMBERS OF MULTINATIONAL STUDY OF CULTURAL DISPLAY RULES. (2008) Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, 94, 925–937.
- MAUSS, I. B., COOK, C. L., & GROSS, J. J. (2007) Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology*, 43, 698–711.
- MAUSS, I. B., EVERS, C., WILHELM, F. H., & GROSS, J. J. (2006) How to bite your tongue without blowing your top: implicit evaluation of emotion regulation predicts affective responding to anger provocation. *Personality and Social Psychology Bulletin*, 32, 589–602.
- MIYAMOTO, Y., & MA, X. (2011) Dampening or savoring positive emotions: a dialectical cultural script guides emotion regulation. *Emotion*, 11, 1346–1357.
- NOVIN, S., RIEFFE, C., & MO, T. H. (2010) The role of situational goals and audience on self-reported emotion experience and expression: Dutch and South Korean children compared. *Infant & Child Development*, 19, 406–421.
- RUSTING, C. L., & NOLEN-HOEKSEMA, S. (1998) Regulating responses to anger: effects of rumination and distraction on angry mood. *Journal of Personality and Social Psychology*, 74, 790–803.

- SCHIMMACK, U. (2001) Pleasure, displeasure, and mixed feelings: are semantic opposites mutually exclusive? *Cognition & Emotion*, 15, 87-91.
- SCHIMMACK, U. (2003) Affect measurement in experience sampling research. *Journal of Happiness Studies*, 4, 79-106.
- SCOLLON, C. N., DIENER, E., OISHI, S., & BISWAS-DIENER, R. (2004) Emotion across cultures and methods. *Journal of Cross-Cultural Psychology*, 35, 304-326.
- SILK, J. S., STEINBERG, L., & MORRIS, A. S. (2003) Adolescents' emotion regulation in daily life: links to depressive symptoms and problem behavior. *Child Development*, 74, 1869-1880.
- SILVERS, J. A., MCRAE, K., GABRIELI, J. D. E., GROSS, J. J., REMY, K. A., & OCHSNER, K. N. (2012) Age-related differences in emotional reactivity, regulation and rejection sensitivity in adolescence. *Emotion*, 12, 1235-1247.
- SOTO, J. A., PEREZ, C. R., KIM, Y.-H., LEE, E. A., & MINNICK, M. R. (2011) Is expressive suppression always associated with poorer psychological functioning? A cross-cultural comparison between European Americans and Hong Kong Chinese. *Emotion*, 11, 1450-1455.
- TAMIR, M. (2005) Don't worry, be happy? Neuroticism, trait-consistent affect regulation, and performance. *Journal of Personality and Social Psychology*, 89, 449-461.
- TAMIR, M., CHIU, C.-Y., & GROSS, J. J. (2007) Business or pleasure? Utilitarian versus hedonic consideration in emotion regulation. *Emotion*, 7, 546-554.
- TAMIR, M., MITCHELL, C., & GROSS, J. J. (2008) Hedonic and instrumental motives in anger regulation. *Psychological Science*, 19, 324-328.
- WEBB, T. L., MILES, E., & SHEERAN, P. (2012) Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, 138, 775-808.

Accepted September 13, 2013.

LONGITUDINAL PREDICTIONS OF THE BROODING AND REFLECTION SUBSCALES OF THE JAPANESE RUMINATIVE RESPONSES SCALE FOR DEPRESSION¹

AKIRA HASEGAWA

*Faculty of Human Relations
Tokai Gakuin University, Japan*

MUNENAGA KODA

*Graduate School of Medicine
University of the Ryukyus, Japan*

YOSUKE HATTORI

*Graduate School of Arts and Sciences
The University of Tokyo, Japan*

TSUYOSHI KONDO

*Graduate School of Medicine
University of the Ryukyus, Japan*

JUN KAWAGUCHI

*Department of Psychology
Nagoya University, Japan*

Summary.—The Ruminative Responses Scale (RRS) is a measure of depressive rumination which has two subscales: Brooding and Reflection. This article examines the longitudinal predictions for depression and the test-retest reliability of the Brooding and Reflection of the Japanese RRS. Japanese university students ($N=378$) completed the RRS, the Center for Epidemiologic Studies Depression Scale (CES-D), and the Inventory to Diagnose Depression, Lifetime Version (IDDL) which was modified to assess symptoms experienced in the 8-wk. follow-up period. The standardized betas of the initial Brooding and Reflection subscales for the IDDL scores were significant and positive after controlling for baseline CES-D scores, but those for the CES-D scores at Time 2 were not significant. Longitudinal predictions of Brooding were partially consistent with those of other language versions (significant in almost all studies); however, longitudinal predictions of Reflection were not consistent with those of other language versions (negative in previous studies). The test-retest correlations of both subscales were similar to those obtained in Western countries.

The Ruminative Responses Scale (RRS; Nolen-Hoeksema & Morrow, 1991) is a measure of depressive rumination, which is defined as repetitive and passive thinking about one's symptoms of depression and the possible causes and consequences of these symptoms (Nolen-Hoeksema, 2004). People with current or remitted major depression scored higher on the RRS than those who had never been depressed (Joormann, Dkane, & Gotlib, 2006; Beevers, Rohde, Stice, & Nolen-Hoeksema, 2007). Numerous

¹Address correspondence to Akira Hasegawa, Tokai Gakuin University, East Campus 360, 5-68 Naka-kirino, Kakamigahara City, Gifu, 504-8511, Japan or e-mail (mail-ad.of.hasse@tokaigakuin-u.ac.jp).

studies have shown that higher scores on the RRS total scale are predictive of more severe depression (Nolen-Hoeksema & Morrow, 1991; Butler & Nolen-Hoeksema, 1994), as well as the onset and relapse of major depressive episodes (Nolen-Hoeksema, 2000; Spasojevic & Alloy, 2001). Consistent with these findings, many experimental studies have shown that rumination increases depression. For example, negative mood increased in dysphoric and depressed participants that were asked to focus on the meanings, causes, and consequences of their current feelings for eight minutes (i.e., induced to ruminate), whereas participants with the identical mood status that were induced to distract from negative thoughts or feelings for eight minutes improved their mood (Nolen-Hoeksema & Morrow, 1993). Induced rumination is considered to lead to various outcomes, such as impaired social problem solving (Lyubomirsky & Nolen-Hoeksema, 1995; Lyubomirsky, Tucker, Caldwell, & Berg, 1999; Donaldson & Lam, 2004), negative interpretations about oneself, one's situation, and future events (Lyubomirsky & Nolen-Hoeksema, 1995; Lyubomirsky, *et al.*, 1999; Lavender & Watkins, 2004), frequent retrieval of negative autobiographical memories and increased memory-related distress (Lyubomirsky, Caldwell, & Nolen-Hoeksema, 1998; Williams & Moulds, 2010), overgeneral autobiographical memories (Park, Goodyer, & Teasdale, 2004; Sutherland & Bryant, 2007), and deficits in cognitive control (Philippot & Brutoux, 2008; Whitmer & Gotlib, 2012). These findings suggest that depressive rumination is a vulnerability factor for depression and one of the core cognitive processes in the pathology of depressive disorders.

Although several versions of the RRS, which differ in terms of item composition, have been proposed (Nolen-Hoeksema & Morrow, 1991; Butler & Nolen-Hoeksema, 1994), in recent studies most researchers have used the version of the RRS revised by Treynor, Gonzalez, and Nolen-Hoeksema (2003) with 22 standard items. Because Treynor, *et al.* (2003) suggested that 12 items of the RRS could be confounded with depressive symptoms, these items were considered unsuitable for investigating the relationships between rumination and depression. The remaining 10 items were subjected to an exploratory factor analysis, yielding two factors: Brooding (i.e., "a passive comparison of one's current situation with some unachieved standard") and Reflective pondering or Reflection (i.e., "a purposeful turning inward to engage in cognitive problem solving to alleviate one's depressive symptoms") (Treynor, *et al.*, 2003, p. 256). Both subscales are composed of five items. The validity of the two-factor structure of the RRS has been supported by other researchers (Schoofs, Hermans, & Raes, 2010).

These two factors appear to be distinct subcomponents, since Brooding and Reflection have different longitudinal predictions for depression. Treynor, *et al.* (2003) conducted a longitudinal study over one year with a

community sample. They assessed Brooding and Reflection, and depression as measured with the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) at baseline and 1 yr. later. They demonstrated that while the standardized beta of Brooding for depression at follow-up was positive ($\beta = .15$, $p < .001$) after controlling for baseline depression scores, that of Reflection was negative ($\beta = -.05$, $p < .05$). This pattern, as noted by the authors, suggests that Brooding may represent a maladaptive aspect of rumination and Reflection may represent an adaptive aspect. In the subsequent studies, Schoofs, *et al.* (2010) conducted a longitudinal study over 6 mo. with undergraduate students, and Pearson, Watkins, & Mullan (2010) conducted a 6-mo. study with currently depressed, previously depressed, and never depressed individuals. They assessed Brooding, Reflection, and depression as measured with the BDI-II (Beck, Steer, & Brown, 1996) at baseline and follow-up tests. Both studies showed that the standardized betas of Brooding for depression at a 6-mo. follow-up were positive ($\beta = .18$, $p < .001$; Schoofs, *et al.*, 2010) after controlling for baseline depression scores, and that of Reflection were not statistically significant ($\beta = -.01$, ns; Schoofs, *et al.*, 2010).² Raes, Smets, Nelis, and Schoofs (2012) conducted two longitudinal studies. The first study that was conducted with secondary school students indicated that the standardized betas of Brooding and Reflection scores for depression assessed by the Depression subscale of the Depression Anxiety Stress Scale (Lovibond & Lovibond, 1995) at 3-mo. follow-up were not significant ($\beta s = .11$ and $-.06$, respectively, ns). In their other study, conducted with university students, the standardized betas of the Brooding score for the Affective and Cognitive subscale scores of the BDI-II at 5-mo. follow-up were positive and significant ($\beta s = .14$ and $.11$, respectively, $ps < .05$), after controlling for the baseline scores on each scale. The standardized beta of the Brooding score for the Somatic subscale score of the BDI-II was not statistically significant ($\beta = .09$, ns). Moreover, the standardized betas of the Reflection score for the BDI-II subscale scores were not significant (range of $\beta s =$ from $-.05$ to $-.02$, ns). The overall results of their studies, with the exception of the secondary school student sample in Raes, *et al.*, (2012) suggested that the Brooding of the RRS assesses a maladaptive aspect of rumination, whereas the Reflection assesses an adaptive or relatively less maladaptive aspect of rumination, at least with respect to the longitudinal prediction

²Pearson, *et al.* (2010) stated that "Time 1 brooding was a significant predictor of Time 2 depressive symptoms" (p. 970); however, the standardized beta of Brooding obtained in their study was .03. It is possible that this beta value is not accurate, because the value does not reach significance in their sample size ($n = 92$), even though the t value was high enough to reach significance ($t = 2.38$). Therefore, we have not discussed the standardized beta of Brooding and Reflection obtained by Pearson, *et al.* (2010) in detail.

of depression.^{3, 4} This notion is consistent with the findings that Brooding is generally related to pathological cognitive and behavioral factors including attentional bias, overgeneral autobiographical memory, concern over mistakes in perfectionism, and passive coping responses, whereas Reflection typically is not (e.g., Joormann, *et al.*, 2006; Harris, Pepper, & Maack, 2008; Debeer, Hermans, & Raes, 2009; Marroquin, Fontes, Scilletta, & Miranda, 2010).⁵

Hasegawa (2013) translated the 22-item version of the RRS based on the item composition of Treynor, *et al.* (2003) into Japanese, and examined the two-factor model of the RRS with a sample of 299 university students (124 women, 162 men; *M* age=20.4 yr., *SD*=2.3; sex and age data were missing for 13 participants). Confirmatory factor analysis indicated that the two-factor model of the RRS showed a moderate to good fit to the data (CFI=.92, NFI=.89, NNFI=.89, and RMSEA=.08). The fit indices of the two-factor model were better than those of a unidimensional model in which the 10 items assessing Brooding and Reflection comprised one factor. In addition, the Japanese version of RRS has good construct validity. Hasegawa (2013) also demonstrated that Brooding was correlated more strongly with current depression, assessed with the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) than was Reflection ($r=.55$ for Brooding; $r=.36$ for Reflection). However, the longitudinal predictions of the Japanese RRS subscales have not yet been examined. This examination would also indicate whether the Japanese version of the RRS has similar psychometric properties as other language versions. Because the RRS has been utilized in numerous studies, including clinical outcome studies, the standardized betas of Brooding and Reflection subscales of the Japanese version of the RRS for depression assessed after an interval of a few months is important for research on depression conducted in Japan, and valuable for the comparison of findings obtained in Japan with that of other countries. For the same reasons, it is important to accumulate data on the reliability of the Japanese version of the RRS.

³There were other studies comparing predictive powers of Brooding and Reflection for future depression. However, they examined this after controlling for other cognitive variables related to rumination, so no direct comparison between the present study and these other studies is possible.

⁴With the exception of the sample of secondary school students in Raes, *et al.* (2012), showing that Brooding was not significantly related to depression at follow-up, longitudinal studies have utilized the BDI or BDI-II as dependent variables. It is possible that the longitudinal relationship between Brooding and depression depends on the depression measure that was utilized (see Discussion section for a detailed analysis).

⁵It is known that Reflection has some detrimental consequences, including suicidal ideation (Miranda & Nolen-Hoeksema, 2007; Surrenre, Miranda, Marroquin, & Chan, 2009). With current knowledge, we cannot fully explain why Reflection decreases, or is unrelated to depression, but simultaneously leads to increased suicidal ideation. Therefore, in the present study, we focused mainly on the relationship between rumination and depression.

The first purpose of the present study is to examine whether Brooding is associated positively with future depression, and whether Reflection is associated negatively with depression, as had been shown in previous studies conducted in Western countries. Also, the standardized beta of the total score on the Japanese RRS for depression assessed 8 wk. later was examined, after controlling for baseline depression scores, since the total scores, as well as RRS subscale scores, had been calculated and analyzed in many of the previous research studies.

The second purpose is to examine test-retest reliability of the Japanese RRS with a large sample. To our knowledge, only a few studies have examined the test-retest reliability of the Brooding and Reflection subscales of the RRS (see Table 1). In studies conducted in Western countries, the range of the test-retest correlations has varied from .53 to .74 for the Brooding and from .35 to .60 for the Reflection. The test-retest correlations of both subscales were generally in the moderate range ($.40 \leq r < .70$), with the exception of the low correlation ($r = .35$) for Reflection reported by Zetsche and Joormann (2011). These results are consistent with the general idea that response styles, as assessed with the Brooding and Reflection subscales, represent relatively trait-like characteristics, but with the possibility of changing according to depressive states (see Beevers, *et al.*, 2007). Hasegawa (2013) examined the test-retest correlations for the Brooding and Reflection subscales of the Japanese RRS, and found that these correlations for each subscale are similar to versions in other languages. However, because the sample size in Hasegawa (2013) was small ($n = 56$), in the

TABLE 1
SUMMARY OF TEST-RETEST RELIABILITY ESTIMATES FOR THE RUMINATIVE RESPONSE SCALE'S SUBSCALES

Study	Sample	<i>n</i>	Interval		<i>r</i>
Treyner, Gonzalez, & Nolen-Hoeksema (2003)	Community sample	1130	1 yr.	Brooding	.62
				Reflection	.60
Pearson, Watkins, & Mul-lan (2011)	Patients and Community sample	92	6 mo.	Brooding	.60
				Reflection	.56
Zetsche & Joormann (2011)	University students	40	6 mo.	Brooding	.74
				Reflection	.35
Raes, Smets, Nelis, & Schoofs (2012)	Secondary school students	143	3 mo.	Brooding	.53
				Reflection	.52
Raes, Smets, Nelis, & Schoofs (2012)	University students	344	5 mo.	Brooding	.57
				Reflection	.47
Hasegawa (2013)	University students	56	6 wk.	Brooding	.59
				Reflection	.72
The present study	University students	378	8 wk.	Brooding	.60
				Reflection	.58

present study the test-retest reliability of the RRS subscales was examined with a larger sample.

To assess whether Brooding and Reflection were distinct constructs, Treynor, *et al.* (2003) conducted multiple regression analyses and found that variances of scores on the Brooding and Reflection subscales at follow-up were best predicted by their own initial subscale scores, even after controlling for the other RRS subscale and BDI scores at baseline. These results indicated that Brooding and Reflection are distinct constructs. The present study tries to replicate these findings with a Japanese sample.

Hypothesis 1. Brooding will be positively associated, and Reflection negatively associated, with future depression scores.

Hypothesis 2. The test-retest reliability of the Brooding and Reflection scales in the Japanese version will be in the moderate range.

Hypothesis 3. Brooding and Reflection subscale scores at follow-up will be best predicted by their own initial subscale scores, even after controlling for the other RRS subscale and depression scores at baseline.

METHOD

Participants

At the beginning of the academic year, 437 undergraduate students (241 women) at the Nagoya University, Tokai Gakuin University, and University of the Ryukyus participated in a questionnaire study (Time 1). Participants were recruited in their classes, mainly psychology classes, and completed the questionnaires there. The mean age of the participants who answered the questionnaire batteries at Time 1 was 19.3 yr. ($SD=2.8$). Eight weeks later (Time 2), the same questionnaire battery was administered to 378 of these students (215 women). The mean age of the participants who completed both questionnaire batteries was 19.3 yr. ($SD=2.9$) at Time 1. To verify whether there were differences between the completed and the dropout participants on age and the questionnaire measures, the means for the completed participants and the drop-out participants were compared. There were no differences at Time 1 using independent t tests ($t_{435} < 1.33$, $p > .18$).

Measures

Ruminative Responses Scale (RRS; Treynor, *et al.*, 2003).—The RRS includes 22 items, each rated on a 4-point rating scale with anchors 1: Almost never and 4: Almost always. The RRS contains five items comprising the Brooding, five items comprising the Reflection, and 12 depression-related items. Brooding and Reflection subscale scores and the total RRS score were com-

puted. Only a few studies have utilized the depression-related items of the RRS as a unitary subscale; therefore, in the current study, these items were not considered a single subscale. Since the development of the RRS (Nolen-Hoeksema & Morrow, 1991), numerous researchers have summed the scores of all the RRS items and analyzed the total RRS score, as reviewed by Nolen-Hoeksema, Wisco, and Lyubomirsky (2008). In line with such previous studies, the total score for the 22 items of the RRS was analyzed in the current study. Adequate psychometric properties were reported including good internal consistency, moderate test-retest reliability, and evidence for construct validity for the total score and subscale scores (Treyner, *et al.*, 2003; Schoofs, *et al.*, 2010). The Japanese translation by Hasegawa (2013) was used, for which good reliability and validity were reported. Cronbach's alphas of the RRS total scale and subscales, and the other scales utilized in this study, are displayed in Table 2.

Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977).—The CES-D is a self-report scale designed to measure depressive symptomatology in the general population. The scale includes 20 items, each rated on a 4-point rating scale with anchors 0: Rarely or none of the time and 3: Most or all of the time. Participants indicated how they were feeling over the last week. Shima, Shikano, Kitamura, and Asai (1985) translated the CES-D into Japanese, and confirmed adequate reliability and validity of the scale.

Inventory to Diagnose Depression, Lifetime Version (IDDL; Zimmerman & Coryell, 1987).—The IDDL is a self-report scale designed to assess the severity of the worst lifetime depressive episode. The scale includes 22 items, each rated on a 5-point rating scale (0 to 4). As an example, the IDDL item for insomnia is: "0: I was not sleeping less than normal." "1: I occasionally had slight difficulty sleeping." "2: I clearly didn't sleep as well as usual." "3: I slept about half my normal amount of time." "4: I slept less than two hours per night." Uehara, Sato, Sakado, and Sato (1995) translated the IDDL into Japanese. The Japanese version has adequate reliability and validity (Uehara, *et al.*, 1995; Sakado, Sato, Uehara, Sato, & Kameda, 1996; Sato, Uehara, Sakado, Sato, Nishioka, & Kasahara, 1996). In the original instructions, participants were instructed to recall their worst lifetime episode of depression. However, this scale was used here to assess the severity of the worst episode of depression experienced by participants in the period from Time 1 through Time 2, and the instructions were modified accordingly. Modification of the instruction did not seem to increase the difficulty of rating, because participants merely responded to each item while recalling the worst episode of depression experienced in the follow-up period, instead of recalling the worst lifetime episode of depression. Participants only filled in the modified IDDL at Time 2.

Procedure

This study was approved by the Ethics Committee of Tokai Gakuin University.

Prior to their participation, all the students were given a complete description of the study, and participants provided written informed consent. They completed the RRS and CES-D at Time 1, and the RRS, CES-D, and IDDL 8 wk. later (Time 2). The 8-wk. interval was adopted so that the study could be completed in one semester for practical reasons to minimize attrition. The rate of attrition in this study was about 14%. Participants who completed both questionnaire batteries received a book card worth 500 yen (approximately \$5 U.S.).

Analysis

Sex differences for scores on the RRS total scale and subscales, CES-D, and IDDL were examined with a *t* test. Zero-order Pearson correlations were computed between each measure to examine the relationships between the total RRS score and subscale scores, as well as with measures of depression, and to examine the test-retest reliability of the RRS total scale and subscales. In line with Schoofs, *et al.* (2010) and Hasegawa (2013), *t* tests were conducted to examine statistically significant differences in the correlations between Brooding and Reflection subscale scores and measures of depression at identical points in time (i.e., the CES-D at the present study). In addition, to re-examine the factor structure of the two-factor model of the Japanese RRS, Confirmatory Factor Analysis (CFA) was conducted with maximum likelihood estimation using data at Time 1 from the participants who completed the two assessments ($N=378$). The factor structure of the two-factor model of the Japanese RRS was re-examined, because the fit indices of the model support the psychometric properties of the Japanese translation. To examine the longitudinal associations of the Brooding score and the Reflection score for depression measured after 8 wk., multiple regression analyses were conducted with the Brooding score, the Reflection score, and the CES-D score at Time 1 as independent variables and the CES-D score at Time 2 and the IDDL score as dependent variables. Multiple regression analyses were also conducted with the total RRS score and CES-D score at Time 1 as independent variables and CES-D score at Time 2 and IDDL score as dependent variables. The CES-D score at Time 1 was included in the model for the IDDL score because to examine longitudinal predictions of rumination measures for depression, baseline depression scores should be controlled. Moreover, to assess if the Brooding and Reflection subscale scores at follow-up were best predicted by their own initial subscale scores, multiple regression analyses were conducted with the Brooding score, Reflection score, and CES-D score at Time 1 as independent variables and Brooding score and Reflection score at Time 2 as dependent variables.

TABLE 2
DESCRIPTIVE STATISTICS FOR EACH SCALE AT TIME 1 AND TIME 2 FOR THE TOTAL SAMPLE, BY SEX,
WITH SEX DIFFERENCES FOR EACH SCALE

Scale	Total Sample (N=378)					Men (n=163)		Women (n=215)		t
	M	SD	α	Skew.	Kurt.	M	SD	M	SD	
1. Brooding, Time 1	11.33	3.71	.77	.31	-.53	11.03	3.73	11.56	3.69	1.37
2. Brooding, Time 2	10.95	3.61	.79	.20	-.59	10.49	3.51	11.30	3.66	2.16*
3. Reflection, Time 1	8.97	3.13	.72	.86	.55	8.76	2.93	9.13	3.27	1.13
4. Reflection, Time 2	8.80	2.99	.69	.88	.49	8.57	2.77	8.98	3.13	1.32
5. RRS Total, Time 1	44.04	12.13	.91	.38	-.42	43.39	12.06	44.53	12.19	0.91
6. RRS Total, Time 2	43.70	12.60	.92	.26	-.66	42.25	12.26	44.80	12.77	1.95
7. CES-D, Time 1	13.99	8.52	.84	.97	1.10	13.22	8.00	14.58	8.87	1.53
8. CES-D, Time 2	14.71	9.04	.86	.97	.74	14.26	8.98	14.96	9.09	0.74
9. IDDL	15.07	10.78	.90	.71	.14	13.92	11.55	16.79	10.92	2.46*

Note.—CES-D=Center for Epidemiologic Studies Depression Scale; RRS=Ruminative Responses Scale; IDDL=Inventory to Diagnose Depression, Lifetime Version; Skew.=skewness; Kurt.=kurtosis. * $p<.05$.

RESULTS

Descriptive statistics for each scale are displayed in Table 2. There were significant sex differences for Brooding at Time 2 and for the IDDL. All of the analyses described below were repeated after controlling for sex, and similar results were observed. Therefore, the analyses are not described after controlling for sex, since sex was not a variable of interest.⁶

Confirmatory Factor Analysis was conducted to confirm the factor structure of the Japanese RRS (i.e., Brooding and Reflection). In line with Treynor, *et al.* (2003) and other previous studies (e.g., Schoofs, *et al.*, 2010; Hasegawa, 2013), the 12 depression-related items were removed and the 10 items that assessed Brooding and Reflection were subjected to factor analysis. Goodness-of fit indices for the model were examined using

⁶Depressive rumination has been proposed as an explanation of the increased likelihood of depression in women than in men, and this notion has been supported by a number of studies (see Nolen-Hoeksema, *et al.*, 2008 for review). However, a previous study conducted in Japan by Sakamoto, Kambara, and Takano (2001), using a different version of the RRS from that later revised by Treynor, *et al.* (2003), did not show significant sex differences in RRS scores. Also, re-analysis of the data by Hasegawa (2013) has indicated that there are no sex differences in the total scale and subscales scores of the Japanese RRS.

the Comparative Fit Index (CFI), the Normed Fit Index (NFI), the Non-Normed Fit Index (NNFI), the Root Mean Square Error of Approximation (RMSEA), and the consistent version of Akaike's Information Criterion (CAIC). CFI, NFI, NNFI, and RMSEA values range from 0 to 1. CFI, NFI, and NNFI values above .90 represent a good model fit, whereas RMSEA values of less than .08 demonstrate an acceptable model fit. Parsimony was assessed using the CAIC, with smaller values representing a better fit. The resulting fit indices indicated that a two-factor model of the RRS showed a moderate fit to the data (CFI = .87, NFI = .84, NNFI = .83, RMSEA = .098, & CAIC = 303.89). Moreover, the fit indices of the two-factor model were better than those for a unidimensional model in which the 10 items of Brooding and Reflection comprised one factor (CFI = .77, NFI = .75, NNFI = .71, RMSEA = .129, & CAIC = 392.78).

Correlations between each measure are displayed in Table 3. The correlation between Brooding and Reflection was .48 at Time 1 and .49 at Time 2. These coefficients were slightly higher than those obtained by Treynor, *et al.* (2003; $r_s = .37-.42$) and Schoofs, *et al.* (2010; $r_s = .33-.42$), but lower than that obtained by Pearson, *et al.* (2010; $r = .72$). The Brooding score was correlated with the CES-D score assessed simultaneously at Times 1 and 2 ($r_s = .50$ and .46, respectively). The Reflection subscale score, compared to the Brooding score, was correlated weakly with the CES-D score assessed simultaneously, at Times 1 and 2 ($r_s = .30$ and .29, respectively). Student *t* tests were conducted to examine statistically significant differences in correlations between Brooding and Reflection subscale scores with the CES-D score at identical time points. Results indicated that the correlation between Brooding score and CES-D score at Time 1 was significantly stronger than

TABLE 3
CORRELATIONS BETWEEN EACH MEASURE

	1	2	3	4	5	6	7	8
1. Brooding, Time 1								
2. Brooding, Time 2	.60							
3. Reflection, Time 1	.48	.35						
4. Reflection, Time 2	.34	.49	.58					
5. RRS Total, Time 1	.86	.59	.74	.49				
6. RRS Total, Time 2	.58	.88	.47	.73	.66			
7. CES-D, Time 1	.50	.37	.30	.25	.55	.46		
8. CES-D, Time 2	.39	.46	.25	.29	.44	.55	.67	
9. IDDL	.45	.54	.35	.46	.52	.63	.54	.65

Note.—CES-D=Center for Epidemiologic Studies Depression Scale; RRS=Ruminative Responses Scale; IDDL=Inventory to Diagnose Depression, Lifetime Version. All correlations statistically significant at $p < .001$.

that with Reflection subscale score at Time 1 ($t_{375} = 4.39, p < .001$). Also, the correlation between Brooding score and CES-D score at Time 2 was significantly stronger than that with the Reflection score at Time 2 ($t_{375} = 3.68, p < .001$). These findings corroborated the results reported in previous studies (Schoofs, *et al.*, 2010; Hasegawa, 2013).

Next, the longitudinal relationships between the RRS subscales at Time 1 and scores for depression assessed at Time 2 were examined. Brooding at Time 1 was positively correlated with CES-D at Time 2 and IDDL ($r_s = .39$ and $.45$, respectively, $ps < .001$), and Reflection at Time 1 was positively correlated with both variables ($r_s = .25$ and $.35$, respectively, $ps < .001$). A multiple regression analysis was conducted with Brooding, Reflection, and CES-D at Time 1 as independent variables and CES-D at Time 2 as a dependent variable. Only CES-D at Time 1 predicted CES-D at Time 2, and Brooding and Reflection were not significant. The independent variables explained 46.5% of the variance of CES-D scores at Time 2. A regression analysis was conducted with the RRS total score and CES-D score at Time 1 as independent variables and CES-D score at Time 2 as the dependent variable. This showed that both variables were significantly related to CES-D scores at Time 2. The independent variables explained 46.8% of the variance in CES-D scores at Time 2 (see Table 4 for the detailed results of two regression analyses).

Next, a multiple regression analysis was conducted with Brooding, Reflection, and CES-D at Time 1 as independent variables and IDDL as a dependent variable. All variables were significantly related to IDDL. The independent variables explained 36.2% of the variance in IDDL scores. A regression analysis was conducted with RRS Total score and CES-D scores at Time 1 as independent variables and IDDL scores as a dependent variable. This showed that both variables were significantly related to IDDL scores. The independent variables explained 36.8% of the variance of IDDL scores (see Table 4 for the detailed results of two regression analyses).

The test-retest correlations for Brooding, Reflection, and the total score of the RRS were $.60$, $.58$, $.66$, respectively (all $ps < .001$; see Table 3). To assess whether Brooding and Reflection are distinct constructs, we used the analyses described in Treynor, *et al.* (2003). A multiple regression analysis was run to examine whether Brooding and Reflection at Time 2 were each best predicted by their own scores at Time 1, even after controlling for the other RRS subscale and CES-D scores at Time 1. Multiple regression analysis with Brooding at Time 2 as the dependent variable showed that Brooding and CES-D scores at Time 1 were significantly related to Brooding at Time 2. The standardized beta of Reflection was not significant. Also, multiple regression analysis with Reflection at Time 2 as the dependent variable

TABLE 4
RESULTS OF MULTIPLE REGRESSION ANALYSES

	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	<i>R</i> ²	Adj <i>R</i> ²	<i>F</i>	<i>p</i>
Dependent variable: CES-D, Time 2									
Brooding, Time 1	.12	.11	.05	1.04	.29				
Reflection, Time 1	.10	.12	.03	0.84	.40				
CES-D, Time 1	.68	.04	.64	14.65	<.001	.46	.46	108.37	<.001
Dependent variable: CES-D, Time 2									
RRS Total, Time 1	.07	.03	.10	2.33	.02				
CES-D, Time 1	.65	.04	.62	13.72	<.001	.46	.46	165.26	<.001
Dependent variable: IDDL									
Brooding, Time 1	.55	.15	.18	3.53	<.001				
Reflection, Time 1	.51	.17	.14	3.03	.003				
CES-D, Time 1	.54	.06	.41	8.60	<.001	.36	.35	70.62	<.001
Dependent variable: IDDL									
RRS Total, Time 1	.29	.04	.31	6.33	<.001				
CES-D, Time 1	.49	.06	.37	7.61	<.001	.36	.36	109.20	<.001
Dependent variable: Brooding, Time 2									
Brooding, Time 1	.51	.05	.53	10.38	<.001				
Reflection, Time 1	.07	.05	.06	1.43	.15				
CES-D, Time 1	.03	.02	.09	1.95	.05	.38	.37	77.07	<.001
Dependent variable: Reflection, Time 2									
Brooding, Time 1	.03	.04	.04	0.89	.37				
Reflection, Time 1	.51	.04	.54	11.41	<.001				
CES-D, Time 1	.02	.01	.06	1.40	.16	.35	.34	67.82	<.001

Note.—CES-D=Center for Epidemiologic Studies Depression Scale; RRS=Ruminative Responses Scale; IDDL=Inventory to Diagnose Depression, Lifetime Version.

showed that only Reflection at Time 1 was significantly related to Reflection at Time 2, and Brooding and CES-D scores at Time 1 were not significantly related. This showed that Brooding and Reflection at Time 2 were each best predicted by their own subscale scores at Time 1, even after controlling for the other RRS subscale and CES-D scores at Time 1 (see Table 4 for the detailed results of two regression analyses).

DISCUSSION

The first purpose of the present study was to examine whether Brooding was associated positively with future depression scores, and whether Reflection was associated negatively with depression, as had been found in previous studies conducted in Western countries. A multiple regression analysis showed that Brooding at Time 1 was not significantly related to CES-D scores at Time 2. However, the same regression analysis showed

that Brooding at Time 1 was significantly related to IDDL scores. Therefore, the hypothesis that Brooding was related longitudinally to depression was supported by the analysis with IDDL as a dependent variable.

Why did this inconsistency between predictions of CES-D and IDDL scores occur? In the regression analysis with CES-D at Time 2 as a dependent variable, the same scale score was entered as an independent variable. It could be argued that Brooding at Time 1 did not explain the residual variance of CES-D scores at Time 2 after partialling out the influence of CES-D at Time 1, because CES-D scores at Time 1 were more strongly related to CES-D scores at Time 2, than were IDDL scores assessed at Time 2. Since the correlation between CES-D at Times 1 and 2 was slightly higher than that of CES-D at Time 1 and IDDL ($r_s = .67, .54$, respectively), this explanation seems reasonable. To examine longitudinal predictions of rumination measures for depression, baseline depression should be controlled. Future study should re-examine the present findings with depression measures unified at Times 1 and 2. However, because correlation between Brooding at Time 1 and IDDL ($r = .45$) was slightly higher than that between Brooding at Time 1 and CES-D at Time 2 ($r = .39$), it is possible that Brooding was associated more strongly with the construct assessed by IDDL. Therefore, there are two other possible explanations of the inconsistent longitudinal predictions of Brooding for CES-D at Time 2 and IDDL.

Firstly, the CES-D at Time 2 assesses depressive symptoms at Time 2, while the IDDL assesses the severity of the worst episodes of depressive symptoms experienced in the period from Times 1 through 2. Theoretically, depressive rumination (especially the brooding component of rumination) increases the severity and duration of depression (see Nolen-Hoeksema, 1991). Some participants experienced their worst episodes of depressive symptoms in the follow-up period at Time 2, whereas others experienced the worst episodes in the days of the follow-up period other than Time 2. Therefore, it is plausible that the Brooding score was longitudinally related to the worst episode of depressive symptoms experienced in the period from Times 1 through 2 (assessed with the IDDL), but that the Brooding score was not related to depressive symptoms at Time 2 (assessed with the CES-D). However, the seminal paper of Treynor, *et al.* (2003) and subsequent studies (Pearson, *et al.*, 2010; Schoofs, *et al.*, 2010; Raes, *et al.*, 2012, undergraduate sample) showed that Brooding was significantly related to future BDI or BDI-II scores, which assessed depressive symptoms at Time 2. Therefore, it is possible that Brooding is significantly related to depressive symptoms at follow-up times, as assessed with measures other than the CES-D.

The second possibility involves the constructs each of the depression scales assesses. The IDDL, BDI, and BDI-II are composed of items repre-

senting the symptom criteria for major depressive episodes. On the other hand, the items of the CES-D do not cover all of the symptoms of major depression (e.g., suicidal ideation and attempts) and include four reverse-scored items. Positive emotions and cognitions that are assessed with the reverse-scored items have different functions than negative emotions and cognitions, which are the main symptoms of depression (e.g., Brown, Chorpita, & Barlow, 1998). Therefore, the CES-D, which was developed for use with general populations, may assess depressive mood states, but not clinical depressive symptoms in the same way as more clinical instruments. Previous studies have demonstrated that Brooding was significantly related to future depression utilizing the BDI and BDI-II (Treyner, *et al.*, 2003; Pearson, *et al.*, 2010; Schoofs, *et al.*, 2010; Raes, *et al.*, 2012, undergraduate sample). The present study also showed that Brooding was significantly related to IDDL scores, which represent the symptom criteria for major depression, similar to the BDI-II. Therefore, it is plausible that Brooding is associated with symptoms of clinical depression that are listed as the A criteria for major depressive episodes (DSM-IV-TR; American Psychiatric Association, 2000). In line with this suggestion, Raes, *et al.* (2012) conducted a regression analysis, on the secondary school student sample using the Depression subscale score of the Depression Anxiety Stress Scale as the dependent variable, and found that Brooding was not significantly related to depression measured 3 mo. later. This could be because the DASS, similar to the CES-D, does not cover all the symptoms of major depression (e.g., suicidal ideation and attempts, and indecisiveness) and may only assess depressive mood states.

Examination of the possibility that Brooding is associated with symptoms of clinical depression is important to evaluate whether the Brooding scale of the Japanese RRS has similar psychometric properties as versions in other languages. Future research should utilize the BDI-II as a dependent variable in a longitudinal study to compare the psychometric properties of the Japanese Brooding scale and other language versions.

In terms of Reflection, similar longitudinal predictions were found as for Brooding: Reflection did not predict CES-D scores at Time 2 but predicted IDDL scores. Importantly, the association between Reflection and IDDL scores was positive, indicating that people with frequent Reflection also experience increased future depression. Treyner, *et al.* (2003) found that the standardized beta of Reflection for depression assessed at 1 yr. follow-up was negative, which is inconsistent with the present findings. Subsequent studies have shown that the standardized betas of Reflection for depression assessed after 3 to 6 mo. were negative, although not significant (range $-.06$ to $-.01$; Schoofs, *et al.*, 2010; Raes, *et al.*, 2012). It is possible that the inconsistent findings between the present study and

previous studies are due to cultural differences between participants in this and previous studies. Treynor, *et al.* (2003) suggested that Reflection “may eventually be adaptive in reducing negative affect, perhaps because it leads to effective problem solving” (p. 257). In line with this suggestion, Marroquin, *et al.* (2010) showed that Reflection, in comparison to Brooding was significantly correlated with active coping, and weakly correlated with passive coping. The above possibility in combination with the findings of this study suggest that because Reflection leads to active social problem solving in Western populations, their thought patterns could be negatively related to depression, or have a non-significant relationship with depression. In the present study, Reflection was positively related to IDDL scores. This finding together with the above suggestions indicates the possibility that reflective thought patterns in Japanese people seldom lead to active social problem solving and may cause repetitive and negative cognitive processes. However, this possibility has never been investigated. We are currently conducting a study to investigate the relationships between depressive rumination and social problem solving in a Japanese sample. The results of this study may show whether or not Reflection in Japanese people leads to active problem solving. There are other possible explanations of the inconsistency, including the different measures of depression (BDI or IDDL), or simply a failure of replication. However, Reflection is associated with more suicidal ideation (Miranda & Nolen-Hoeksema, 2007; Surrence, *et al.*, 2009), indicating that Reflection has some maladaptive aspects. Therefore, it is necessary to re-examine the psychometric properties of Reflection from the Japanese RRS and other language versions in the future.

In the present study, Brooding and Reflection showed similar longitudinal predictions for depression. Therefore, it is possible to argue that Brooding and Reflection subscales of the Japanese RRS were not distinct subcomponents. The correlations between Brooding and Reflection in the present study were slightly higher than those obtained in Treynor, *et al.* (2003) and Schoofs, *et al.* (2010). Also, CFA of the present data indicated that the two-factor model of the RRS showed a moderate fit to the data. However, the fit indices of the two-factor model were better than those of a unidimensional model, indicating that the 10 items assessing Brooding and Reflection consists of two subscales, rather than one factor. Moreover, the fit indices of the present study were similar to those of the same model obtained in Schoofs, *et al.* (2010; CFI = .87, NFI = .85, RMSEA = .14).⁷ The correlations between Brooding and Reflection in the present study were lower than that obtained by Pearson, *et al.* (2010), indicating that the indepen-

⁷Schoofs, *et al.* (2010) indicated that the modified model, with error covariances between two pairs of items allowed, showed acceptable fit to the data (CFI = .96, NFI = .94, RMSEA = .08).

dence of RRS subscales was no more problematic in the Japanese translation than in other language versions of the RRS. In addition, the present study showed that Brooding and Reflection at 8 wk. were best predicted by the prior score of each subscale, even after controlling for the other RRS subscale score and baseline depression level. This is consistent with the formulation of Treynor, *et al.* (2003) that Brooding and Reflection are distinct components. More studies are needed to examine the factor structure and discrimination of the two subscales of the Japanese RRS.

The present study showed that the RRS Total score was significantly related to depression, assessed with the CES-D and IDDL at follow-up. The reason for the discrepancy between the significant longitudinal association of the total RRS score for CES-D and the non-significant association of Brooding score could be due to the different number of items in the two scales. Cronbach's alpha of the RRS total scale at Time 1 (.91) was higher than that of Brooding at Time 1 (.77). It is plausible that the lower internal consistency of the Brooding weakened its longitudinal association with depression. It could be assumed that because the total RRS scale has 12 depression-related items, these items simply reflect the severity of depressive symptoms, rather than ruminative tendencies. However, since the longitudinal association of the RRS Total scale was examined after controlling for baseline depression scores, the significant longitudinal association of the RRS could not be simply the result of depression-related items of the RRS. Moreover, it has been reported that people with major depression in remission scored higher on the RRS Total scale than those who had never been depressed (Joormann, *et al.*, 2006; Beevers, *et al.*, 2007), indicating that the RRS Total scale score was not contaminated by depressive symptoms at the same time point. Therefore, it is concluded that the Japanese version of the total RRS scale is a valid instrument for assessing vulnerability to depression and for research and clinical purposes.

The second purpose of the present study was to examine the test-retest reliability of the Japanese RRS with a relatively large sample. The test-retest correlations of both subscales were in the moderate range, and similar to the values obtained in previous studies conducted in Japan and Western countries (see Table 1). These results are consistent with the general idea that response styles, as assessed with the Brooding and Reflection subscales, represent relatively trait-like characteristics. In addition, these findings indicate that the reliability of the Japanese Brooding and Reflection subscales, is similar to the other language versions.

Overall, the present study showed that the Brooding and Reflection subscales of the Japanese RRS at Time 1 were not significantly related to CES-D scores at Time 2, but were associated positively with IDDL scores. These longitudinal predictions using each subscale were not completely consistent

with findings from other language versions⁸ of the RRS. The inconsistencies might be due to different depression measures utilized as dependent variables, or cultural differences. Second, CFA with the present data indicated that the two-factor model of the Japanese RRS showed a moderate fit to the data. More information about the factor structure of the Japanese RRS is necessary. Thirdly, at least the total scale of the Japanese RRS can be utilized to assess vulnerability for depression. Fourthly, the subscales of the Japanese RRS have similar test-retest reliability as found with other language versions. Finally, consistent with Treynor, *et al.* (2003), Brooding and Reflection at the follow-up test were best predicted by their own prior scores, indicating that Brooding and Reflection are distinct components. The RRS has been utilized in a number of basic scientific studies and clinical outcome studies. The comparability of the psychometric properties of the scales is important for the use of these scales in research on depression in Japan. The present findings indicate that most aspects of the Japanese RRS are similar to other language versions, while some aspects may be different. This must be considered when the Japanese RRS is utilized and results are interpreted. Also, these consistencies and inconsistencies need to be replicated.

There are some limitations in the present study. The correlations between depression measures (especially CES-D) assessed at Times 1 and 2 were rather high, which might have made the longitudinal association of rumination measures for depression weaker. For practical reasons, an 8-wk. interval between the baseline and follow-up assessments was adopted, but this short interval might have lead to increased correlations between scores on depression measures assessed at Times 1 and 2. It is suggested that future studies should adopt a longer interval between the two assessments. Participants were university students rather than a clinical population. Information about psychometric properties of the Japanese RRS and the longitudinal associations of its subscales for depression in the clinical sample would be valuable for the clinical practice. It is possible that in clinical populations, Brooding would be related positively to the CES-D score assessed after an interval of 8 wk. and Reflection would be related negatively to the future CES-D score, although this is speculative, because only a few studies (e.g., Arditte & Joormann, 2011) have examined the longitudinal relationships between the RRS subscales and depression.⁸ Future studies should re-examine the relationships among Brooding, Reflection, and depression in clinical samples. The data were self-report measures to assess mood status, so future studies should re-

⁸Arditte and Joormann (2011) examined the longitudinal predictions of Brooding and Reflection after controlling for other emotional regulation strategies. Because direct comparison between the present study and Arditte and Joormann (2011) is impossible, we have not described the findings in detail.

examine the relationships using observer-rated scales, such as the Hamilton Rating Scale for Depression.

REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION. (2000) *Diagnostic and statistical manual of mental disorders*. (4th ed., text rev.) Washington, DC: Author.
- ARDITTE, K. A., & JOORMANN, J. (2011) Emotion regulation in depression: reflection predicts recovery from a major depressive episode. *Cognitive Therapy and Research*, 35, 536-543.
- BECK, A. T., STEER, R. A., & BROWN, G. K. (1996) *Manual for the Beck Depression Inventory-II*. San Antonio, TX: The Psychological Corporation.
- BECK, A. T., WARD, C. H., MENDELSON, M., MOCK, J., & ERBAUGH, J. (1961) An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- BEEVERS, C. G., ROHDE, P., STICE, E., & NOLEN-HOEKSEMA, S. (2007) Recovery from major depressive disorder among female adolescents: a prospective test of the scar hypothesis. *Journal of Consulting and Clinical Psychology*, 75, 888-900.
- BROWN, T. A., CHORPITA, B. F., & BARLOW, D. H. (1998) Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, 107, 179-192.
- BUTLER, L. D., & NOLEN-HOEKSEMA, S. (1994) Gender differences in responses to depressed mood in a college sample. *Sex Roles: A Journal of Research*, 30, 331-346.
- DEBEER, E., HERMANS, D., & RAES, F. (2009) Associations between components of rumination and autobiographical memory specificity as measured by a Minimal Instructions Autobiographical Memory Test. *Memory*, 17, 892-903.
- DONALDSON, C., & LAM, D. (2004) Rumination, mood and social problem-solving in major depression. *Psychological Medicine*, 34, 1309-1318.
- HARRIS, P. W., PEPPER, C. M., & MAACK, D. J. (2008) The relationship between maladaptive perfectionism and depressive symptoms: the mediating role of rumination. *Personality and Individual Differences*, 44, 150-160.
- HASEGAWA, A. (2013) Translation and initial validation of the Japanese version of the Ruminative Responses Scale. *Psychological Reports*, 112, 716-726.
- JOORMANN, J., DKANE, M., & GOTLIB, I. H. (2006) Adaptive and maladaptive components of rumination: diagnostic specificity and relation to depressive biases. *Behavior Therapy*, 37, 269-280.
- LAVENDER, A., & WATKINS, E. (2004) Rumination and future thinking in depression. *British Journal of Clinical Psychology*, 43, 129-142.
- LOVIBOND, P. F., & LOVIBOND, S. H. (1995) The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335-343.
- LYUBOMIRSKY, S., CALDWELL, N. D., & NOLEN-HOEKSEMA, S. (1998) Effects of ruminative and distracting responses to depressed mood on retrieval of autobiographical memories. *Journal of Personality and Social Psychology*, 75, 166-177.
- LYUBOMIRSKY, S., & NOLEN-HOEKSEMA, S. (1995) Effects of self-focused rumination on negative thinking and interpersonal problem solving. *Journal of Personality and Social Psychology*, 69, 176-190.

- LYUBOMIRSKY, S., TUCKER, K. L., CALDWELL, N. D., & BERG, K. (1999) Why ruminators are poor problem solvers: clues from the phenomenology of dysphoric rumination. *Journal of Personality and Social Psychology*, 77, 1041-1060.
- MARROQUIN, B. M., FONTES, M., SCILLETTA, A., & MIRANDA, R. (2010) Ruminative subtypes and coping responses: active and passive pathways to depressive symptoms. *Cognition & Emotion*, 24, 1446-1455.
- MIRANDA, R., & NOLEN-HOEKSEMA, S. (2007) Brooding and reflection: rumination predicts suicidal ideation at 1-year follow-up in a community sample. *Behaviour Research and Therapy*, 45, 3088-3095.
- NOLEN-HOEKSEMA, S. (1991) Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, 100, 569-582.
- NOLEN-HOEKSEMA, S. (2000) The role of rumination in depressive disorders and mixed anxiety / depressive symptoms. *Journal of Abnormal Psychology*, 109, 504-511.
- NOLEN-HOEKSEMA, S. (2004) The response styles theory. In C. Papageorgiou & A. Wells (Eds.), *Depressive rumination: nature, theory, and treatment*. West Sussex: John Wiley & Sons. Pp. 107-123.
- NOLEN-HOEKSEMA, S., & MORROW, J. (1991) A prospective study of depression and post-traumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. *Journal of Personality & Social Psychology*, 61, 115-121.
- NOLEN-HOEKSEMA, S., & MORROW, J. (1993) Effects of rumination and distraction on naturally occurring depressed mood. *Cognition & Emotion*, 7, 561-570.
- NOLEN-HOEKSEMA, S., WISCO, B. E., & LYUBOMIRSKY, S. (2008) Rethinking rumination. *Perspectives on Psychological Science*, 3, 400-424.
- PARK, R. J., GOODYER, I. M., & TEASDALE, J. D. (2004) Effects of induced rumination and distraction on mood and overgeneral autobiographical memory in adolescent Major Depressive Disorder and controls. *Journal of Child Psychology and Psychiatry*, 45, 996-1006.
- PEARSON, K. A., WATKINS, E. R., & MULLAN, E. G. (2010) Submissive interpersonal style mediates the effect of brooding on future depressive symptoms. *Behaviour Research and Therapy*, 48, 966-973.
- PEARSON, K. A., WATKINS, E. R., & MULLAN, E. G. (2011) Rejection sensitivity prospectively predicts increased rumination. *Behaviour Research and Therapy*, 49, 597-605.
- PHILIPPOT, P., & BRUTOUX, F. (2008) Induced rumination dampens executive processes in dysphoric young adults. *Journal of Behavior Therapy and Experimental Psychiatry*, 39, 219-227.
- RADLOFF, L. S. (1977) The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- RAES, F., SMETS, J., NELIS, S., & SCHOofs, H. (2012) Dampening of positive affect prospectively predicts depressive symptoms in non-clinical samples. *Cognition & Emotion*, 26, 75-82.
- SAKADO, K., SATO, T., UEHARA, T., SATO, S., & KAMEDA, K. (1996) Discriminant validity of the Inventory to Diagnose Depression, Lifetime Version. *Acta Psychiatrica Scandinavica*, 93, 257-260.
- SAKAMOTO, S., KAMBARA, M., & TANNO, Y. (2001) Response styles and cognitive and affective symptoms of depression. *Personality and Individual Differences*, 31, 1053-1065.
- SATO, T., UEHARA, T., SAKADO, K., SATO, S., NISHIOKA, K., & KASAHARA, Y. (1996) The test-retest reliability of the Inventory to Diagnose Depression, Lifetime Version. *Psychopathology*, 29, 154-158.

- SCHOofs, H., HERMANS, D., & RAES, F. (2010) Brooding and reflection as subtypes of rumination: evidence from confirmatory factor analysis in nonclinical samples using the Dutch Ruminative Responses Scale. *Journal of Psychopathology and Behavioral Assessment*, 32, 609-617.
- SHIMA, S., SHIKANO, T., KITAMURA, T., & ASAI, M. (1985) Atarashi yokuutusei jiko hyouka syakudo ni tuite [A new self-report depression scale]. *Psychiatry*, 27, 717-723. [In Japanese]
- SPASOJEVIC, J., & ALLOY, L. B. (2001) Rumination as a common mechanism relating depressive risk factors to depression. *Emotion*, 1, 25-37.
- SURRENCE, K., MIRANDA, R., MARROQUIN, B. M., & CHAN, S. (2009) Brooding and reflective rumination among suicide attempters: cognitive vulnerability to suicidal ideation. *Behaviour Research and Therapy*, 47, 803-808.
- SUTHERLAND, K., & BRYANT, R. A. (2007) Rumination and overgeneral autobiographical memory. *Behaviour Research and Therapy*, 45, 2407-2416.
- TREYNOR, W., GONZALEZ, R., & NOLEN-HOEKSEMA, S. (2003) Rumination reconsidered: a psychometric analysis. *Cognitive Therapy and Research*, 27, 247-259.
- UEHARA, T., SATO, T., SAKADO, K., & SATO, S. (1995) Nihongo ban Inventory to Diagnose Depression Lifetime Version (IDDL) no sinraisei to datousei no kentou: Jiko kinyuusiki situmonsi ni yoru utubyou syougai kioueki tyousa ni tuite [The reliability and validity of the Japanese version of the Inventory to Diagnose Depression, Lifetime Version (IDDL)]. *Archives of Psychiatric Diagnosis and Clinical Evaluation*, 6, 73-81. [In Japanese]
- WHITMER, A. J., & GOTLIB, I. H. (2012) Switching and backward inhibition in major depressive disorder: the role of rumination. *Journal of Abnormal Psychology*, 121, 570-578.
- WILLIAMS, A. D., & MOULDS, M. L. (2010) The impact of ruminative processing on the experience of self-referent intrusive memories in dysphoria. *Behavior Therapy*, 41, 38-45.
- ZETSCHKE, U., & JOORMANN, J. (2011) Components of interference control predict depressive symptoms and rumination cross-sectionally and at six months follow-up. *Journal of Behavior Therapy and Experimental Psychiatry*, 42, 65-73.
- ZIMMERMAN, M., & CORYELL, W. (1987) The Inventory to Diagnose Depression, Lifetime Version. *Acta Psychiatrica Scandinavica*, 75, 495-499.

Accepted September 27, 2013.

THE STRUCTURE OF PROSOCIAL BEHAVIOR:
A COMMENT ON RICHAUD, MESURADO, AND CORTADA (2013)¹

MARIOLA ŁAGUNA AND PIOTR OLEŚ

Institute of Psychology
The John Paul II Catholic University of Lublin, Poland

Summary.—The question of how many factors are required to explain prosocial behavior in adolescents was examined. Richaud, Mesurado, and Cortada (2013) have tested two alternative models of prosocial behaviors. In this Comment we (1) discuss the theoretical basis for prosocial behavior models in adolescents and (2) propose possible alternative models as suggestions for further studies.

The core problem addressed in the paper by Richaud, Mesurado, and Cortada (2013) is the dimensionality of prosocial behavior. They presented an Argentinean adaptation of the Prosocial Tendencies Measure (PTM; Carlo & Randall, 2002) and tested two alternative models of prosocial behaviors. The first one was a replication of the six-factor model proposed by Carlo and Randall (2002); the second was developed as a result of the analysis of correlations obtained when estimating the first model. In this Comment we will, firstly, discuss the theoretical basis for prosocial behavior models in adolescents and, secondly, propose possible models alternative to that of Carlo and Randall (2002), which might encourage future research on the structure of prosocial behavior.

When specifying a structural model, researchers may aim to: (1) conceptualize the theory under study, a theory-driven procedure (the recommended method of model specification; MacCallum & Austin, 2000) or (2) modify the initial model on the basis of the data (covariance) obtained in the study, a data-driven procedure. The final model accepted by Richaud and colleagues (2013) was constructed using a data-driven rather than theory-driven process. When faced with two alternative models that were not significantly different from each other, the authors decided to accept the reduced model instead of the original one. Three previously separate latent variables were joined together due to high correlations obtained in the first estimated model. Such a process is not free from limitations caused by the sampling effect (MacCallum & Austin, 2000). It cannot be ruled out that in other samples or populations the correlation matrix would show different patterns. Relying on theory rather than solely on data prevents

¹Address correspondence to Mariola Łaguna and Piotr Oleś, Institute of Psychology, The John Paul II Catholic University of Lublin, Al. Raclawickie 14, 20-950 Lublin, Poland or e-mail (laguna@kul.pl; oles@kul.pl).

such possible bias. The absence of a strong theoretical rationale and the existence of support for the original six-factor solution obtained in different samples (e.g., Carlo, Knight, McGinley, Zamboanga, & Jarvis, 2010) argue for the choice of the original model.

The six-dimensional model is based on theory as well as empirical results. While three dimensions of prosocial behaviors—Altruistic, Public, and Anonymous—are relatively independent, three others—Dire, Emotional, and Compliant—appear to be highly correlated and similar, “as they all are expected to be elicited by an intense external demand” (Richaud, *et al.*, 2013, p. 694). Thus, according to the four-dimensional model—a solution that is clear-cut and simple—they constitute one dimension, namely Responsive. On the one hand, these three kinds of prosocial behavior are stimulated by external pressure; on the other hand, they are based on different internal needs (Murray, 1938): Dire on succorance and nurturance, Emotional on affiliation and intraception, and Compliant on deference and acceptance. All these different needs have one thing in common: motivation to help another person. From the developmental point of view, one can postulate that these needs, not fully differentiated in children and pre-adolescents, become more and more differentiated in middle and late adolescents. Therefore, further studies conducted in different age groups may provide empirical support for the six-factor model also in an Argentinean population. To conclude, there are important theoretical arguments in favor of the six-dimensional rather than four-dimensional model of prosocial behavior.

In future research on the structure of prosocial behavior, instead of the four-factor solution, other theoretically possible alternatives to Carlo and Randall’s (2002) model could be proposed. One proposition is to check whether altruism constitutes a subdimension of prosocial behavior or whether it is a separate form of personality tendency. Richaud and colleagues (2013) claimed that “although these concepts are connected, they are also different” (p. 688). It is interesting that even the name of this PTM dimension, Altruism, is worded differently (as a noun) than the names of other dimensions (adjectives). It would be interesting to test whether altruism is a distinct phenomenon, only correlated with forms of prosocial behavior, or whether it is one of the subdimensions of such behavior. The first option should be investigated by testing the model with Altruism as one of the latent variables which is correlated (joined through covariances) with other latent constructs, forming a higher-order latent construct: prosocial behavior. The second option should be investigated by testing the model in which Altruism is one of the lower-order latent factors that together form a common higher-order latent dimension. In this way, two additional models could be developed and tested.

Some dimensions of the PTM could be treated as subdimensions of a higher-order latent factor (instead of being combined into a single factor; Richaud, *et al.*, 2013). A multidimensional model could be proposed with three latent variables (Dire, Compliant, and Emotional types of prosocial behavior), which belong to one higher-order latent construct. This seems more consistent with the idea of Carlo and Randall (2002), who distinguished these three separate constructs. Treating these as responses to three kinds of social situations allows them to be distinguished as different latent variables, together forming a second-order latent factor.

Another theoretically possible model may test whether two PTM scales (Anonymous and Public) belong to a single higher-order factor related to the desire for social approval concerning prosocial action. If anonymous behavior is described as “performed without knowledge of who is helped” and public behavior as conducted “by desire to gain approval and respect of others” (Richaud, *et al.*, 2013, p. 689), they both seem to refer—also positively or negatively—to a common underlying feature or formal aspect of behavior: the extent to which it seeks external appreciation. One could therefore test whether the hierarchical model with a second-order latent factor with two lower-order latent subdimensions (latent indicators of a higher-order factor), Anonymous and Public, would be confirmed by the data.

We recommend conducting a systematic inquiry into the dimensionality of prosocial behavior utilizing Edwards' (2001) method of investigating multidimensional constructs. According to this taxonomy, a construct is multidimensional when it refers to several distinct but related dimensions treated as different theoretical concepts. A multidimensional construct may be superordinate, aggregate, or multivariate in nature (Edwards, 2001), and analyzing it as such allows holistic representations of complex phenomena. This analytical strategy can be used to empirically examine alternative theory-driven hierarchical models of prosocial behavior, such as those proposed in this Comment.

Richaud and colleagues (2013) stressed that different types of prosocial behavior have different personal and situational correlates and different antecedents. Taking this into account, one could expect that the dimensions of the Argentinean version of PTM obtained in the analyses will show their specificity in different patterns of correlations with other variables. For example, different prosocial motivations can be, to different extents, correlated with self-esteem; perceiving oneself as a good and worthy person, one may be more ready at least to declare a desire to help others (Baumeister, Campbell, Krueger, & Vohs, 2003). Such results would be a strong confirmation of the distinctness of these dimensions, i.e., their different causes and effects.

Examining the paper by Richaud and colleagues (2013) as a description of the Argentinean adaptation of the PTM, the psychometric properties of the scale should have been presented (e.g., internal consistency—Cronbach's alpha, and the test-retest stability of scores). It is possible that in different samples of adolescents (e.g., differing in age), the scales would show different psychometric characteristics (e.g., Oleś, Alessandri, Oleś, Bak, Jankowski, Łaguna, *et al.*, 2013). This is an area for further studies to explore before the adaptation of the PTM can be used effectively.

REFERENCES

- BAUMEISTER, R. F., CAMPBELL, J. D., KRUEGER, J. I., & VOHS, K. D. (2003) Does high self-esteem cause better performance, interpersonal success, or healthier lifestyles? *Psychological Science in the Public Interest*, 4(1), 1-44. DOI: 10.1111/1529-1006.01431
- CARLO, G., & RANDALL, B. A. (2002) The development of a measure of prosocial behaviors for late adolescents. *Journal of Youth and Adolescence*, 31(1), 31-44. DOI: 10.1023/A:1014033032440
- CARLO, G., KNIGHT, G. P., MCGINLEY, M., ZAMBOANGA, B. L., & JARVIS, L. (2010) The multidimensionality of prosocial behaviors and evidence of measurement equivalence in Mexican American and European American early adolescents. *Journal of Research on Adolescence*, 20, 334-358. DOI: 10.1111/j.1532-7795.2010.00637.x
- EDWARDS, J. (2001) Multidimensional constructs in organizational behavior research: an integrative analytical framework. *Organizational Research*, 4, 144-192.
- MACCALLUM, R. C., & AUSTIN, J. T. (2000) Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201-226.
- MURRAY, H. A. (1938) *Explorations in personality*. New York: Oxford Univer. Press.
- OLEŚ, P. K., ALESSANDRI, G., OLEŚ, M., BAK, W., JANKOWSKI, T., ŁAGUNA, M., & CAPRARA, G. V. (2013) Positive orientation and generalized self-efficacy. *Studia Psychologica*, 1(55), 47-59.
- RICHAUD, M. C., MESURADO, B., & CORTADA, K. (2013) Analysis of dimensions of prosocial behavior in an Argentinean sample of children. *Psychological Reports: Mental & Physical Health*, 111(3), 687-696. DOI: 10.2466/10.11.17.PR0.111.6.687-696

Accepted October 2, 2013.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Relationships
&
Communications*



THE WORKING ALLIANCE IN GROUP COUNSELING: AN EXPLORATORY STUDY¹

ROSTYSLAW W. ROBAK, KELSEY A. KANGOS, SHEILA H. CHIFFRILLER,
AND PAUL W. GRIFFIN

Pace University

Summary.—The working alliance is a well-defined component of the therapeutic relationship. The present exploratory study examined the development of the working alliance within a group counseling context. The participants ($N = 50$) were students in a graduate program in mental health counseling taking an experiential training course in group dynamics. Participants rated their perceptions of the working alliance at the end of each session in a time-limited training group. Group members' perceptions of their alliance with the group leader and the group as a whole were positively correlated. Outcomes of the group experience were strongly related to the perceived strength of the working alliance as early as the fourth session. Outcomes were also positively correlated with the bonding and agreement-on-goals aspects of the working alliance, but not with the group tasks aspect. The more that group members reported strong working alliances, the more they tended to report that they had self-disclosed in the group.

The therapeutic relationship is recognized as one of the most important factors operating in all approaches to counseling and psychotherapy. Wampold (2007), for instance, stated that communication between therapist and patient is the primary mechanism of change in therapy. The therapeutic relationship creates a climate of cooperation to achieve the clients' goals, therein fostering change. Because the therapeutic alliance (an operationalized term for "relationship") is seen as such a central component to the process of therapy, understanding its effect is important.

Four "therapeutic" factors have been found to be common to all therapy regardless of the therapist's theoretical orientation, the mode of therapy (e.g., individual or group), the "dosage" of therapy (e.g., the frequency of sessions), or the type of problem. These factors are the therapy relationship, client and extra-therapeutic factors (e.g., the passage of time), the theoretical model (and associated techniques), and the placebo, or hope and expectancy of change (Lambert, 1992). The therapeutic relationship has been demonstrated to contribute about 30% of the clients' change (Asay & Lambert, 1999; Duncan, 2010).

Most of the research regarding the working alliance has been done within the context of individual counseling and psychotherapy. Perhaps the most important aspect of the therapeutic relationship is the work-

¹Address correspondence to Rostyslaw Robak, Ph.D., Department of Psychology, Pace University, Pleasantville, New York, 10570 or e-mail (rrobak@pace.edu).

ing alliance that develops between the therapist and the other person in the relationship (i.e., the client). The working alliance is the "strength of the collaborative relationship between client and therapist in therapy" (Luborsky, Barber, Siqueland, & Johnson, 1996; Horvath & Bedi, 2002, p. 41). Although the concept of a working alliance originally derived from psychoanalytic theory, Bordin (1979) defined it in terms that make it universally applicable. Specifically, the three key elements of a working alliance between the therapist and the other person are: the affective bond between the therapist and the other person in the therapeutic relationship, agreement on the goals of the therapy, and collaboration on the tasks of a particular therapeutic endeavor.

The working alliance has been estimated to account for 7% to 17% of the variance in outcomes of therapy (Beutler, Malik, Alimohamed, Harwood, Talebi, & Noble, 2004). A review of several meta-analyses indicates that the correlations between the strength of a working alliance and the therapy outcome range from .22 to .26 (Castonguay, Constantino, & Holtforth, 2006). The low-end result (.22) is from a meta-analysis that coded all reported statistically non-significant results as zero, and thus this lower-end correlation is likely a low estimate (Martin, Garske, & Davis, 2000). In addition, a good therapeutic alliance is strongly associated with decreased rates of dropout from therapy (Arnou, Blasey, Manber, Constantino, Markowitz, Klein, *et al.*, 2007) and lower rates of premature termination (Tracey & Kokotovic, 1989; Tryon & Kane, 1993).

A good working alliance has predicted successful therapeutic outcomes across a range of treatment populations and modalities (Horvath & Symonds, 1991; Martin, *et al.*, 2000). For example, a good working alliance has been associated with positive outcomes in the treatment of alcoholism (Connors, Carroll, DiClemente, Longabaugh, & Donovan, 1997), depression (McBride, Zuroff, Ravitz, Koestner, Moskowitz, Quilty, *et al.*, 2010; Webb, DeRubeis, Amsterdam, Shelton, Hollon, & Dimidjian, 2011), post-traumatic stress disorder (Patterson, Uhlin, & Anderson, 2008; Keller, Zoellner, & Feeny, 2010), and anorexia nervosa (Pereira, 2010).

The working alliance has proven to be remarkably adaptable to a variety of theoretical orientations. It has been used in the context of cognitive-behavioral therapy (d'Iuso, Blake, Fitzpatrick, & Drapeau, 2009), interpersonal therapy for depression (Constantino, Schwaiger, Smith, DeGeorge, McBride, Ravitz, *et al.*, 2010), and psychodynamically-oriented therapy (e.g., Crowe & Grenyer, 2008). It has been shown to be positively correlated with positive outcomes in rehabilitation counseling (e.g., Lustig, Strauser, Weems, Donnell, & Smith, 2003) and in counselor supervision relationships (O'Dell, 2009; Bilodeau, Savard, & Lecomte, 2010). There is little research regarding the working alliance within the context of group counseling and psycho-

therapy. However, a few studies have been done within this setting. These include couples counseling (Knerr, Bartle-Haring, McDowell, Adkins, Delaney, Gangamma, *et al.*, 2011), family therapy (Pereira, Lock, & Oggins, 2006), and group therapy (Taft, Murphy, King, Musser, & DeDeyn, 2003; Bakali, Baldwin, & Lorentzen, 2009).

Group therapy poses an interesting question with regard to working alliance: With whom is the alliance made? Is it with the group leader or with the group as a whole? To date, one study has begun to address this question, albeit indirectly, by asking whether it is the working alliance with the therapist or the factors of group cohesiveness and group climate that predict outcomes (Crowe & Grenyer, 2008). In that study, the alliance with the therapist was not predictive of outcome, but cohesiveness and climate were predictive.

One particularly interesting study pursued the question of how the working alliance is related to self-disclosure in therapy. Kelly has extended her studies of clients' keeping relevant secrets from their therapists by asking whether such lack of self-disclosure is related to outcome in the form of symptom reduction (Kelly & Yuan, 2009). She found that clients who reported keeping a relevant secret (27.7% of the sample) rated their working alliance as being weaker than those who did self-disclose. Those who kept such secrets also showed less symptom reduction.

Many studies have used the ratings of both the therapist and the client in assessing the working alliance. However, Castonguay, *et al.* (2006) have pointed out that therapists' and clients' ratings of the alliance have been found to diverge, especially early in treatment. Furthermore, the clients' perspective is generally more predictive of therapy outcome than the therapists' perspective (Wampold, 2001; Horvath & Bedi, 2002). Therapists should recognize that they may not be able to accurately evaluate the alliance they have with their clients. Interestingly, this research suggests that positive therapeutic outcome is linked with a similarity between therapists' perceptions and client evaluations of the client-therapist alliance at the mid- and late- phases of therapy.

A consensus seems to have emerged that strength of working alliance early in therapy can predict positive outcome (Castonguay, *et al.*, 2006). That point is often reached around the third session in individual therapy (Luborsky, *et al.*, 1996; Horvath & Bedi, 2002). Whether this holds true within group counseling remains an open question. In this study, the development of the working alliance in a time-limited group setting was examined over the course of eight sessions. The relationships between members' perceptions of the working alliance and their perceptions of the outcomes of the group experience for themselves and for others were also investigated.

This was the first of several studies that are currently being conducted. The research questions and hypotheses were:

Research Question. Is there a significant correlation between ratings of the working alliance that group members report with the therapist and with the group as a whole?

Hypothesis 1. Group members' early ratings of the working alliance would be statistically significantly correlated with outcomes, as would ratings at the end of therapy.

Hypothesis 2. The stronger the ratings of working alliance, the more likely the person is to have self-disclosed in the group.

METHOD

Participants

Fifty-four graduate students taking a "Group Dynamics" course as part of their master's program in mental health counseling at a mid-sized university in the New York City area voluntarily participated in the study without compensation. Because four of them did not complete portions of all of the questionnaires, they were dropped from the study. Thus, 50 participants' responses were used in the data analysis; 31 were women, 7 were men, and the rest did not report their sex. Reported ages ranged from 22 to 53 years ($M = 27.2$, $SD = 7.8$) (please note that the "age" question stated that this response was optional and 16 people did not report their age).

Because the questionnaires were collected in small groups of 9 to 11 participants with limited numbers of ethnic minority members in each group, a demographic questionnaire was not included in order to protect participants' anonymity and thus facilitate their openness in responding. However, this course is taken by every student in the program and the program's demographics during the year of data collection were as follows: White (non-Hispanic) 62%, Black (non-Hispanic) 24%, Hispanic 9%, Asian (or Pacific Islander) 2%, non-resident alien 3%, American Indian/Alaskan native 0%.

Procedure

After receiving Institutional Review Board approval, students taking five consecutive sections of a course in group dynamics, beginning in the summer of 2010 and ending in the summer of 2011, were included in the study. All course sections were conducted as a training group of eight sessions. All groups were led by the same professor. Daily follow-up online assignments related to the *in vivo* sessions were a concurrent part of the course. The eight sessions were all two- to two-and-a-half hours long and were structured to simulate the stages of a typical therapy group's life as outlined in *Theory and Practice of Group Counseling* (Corey, 2008). This was accomplished with a combination of structured activities and open dis-

cussions. Structured activities included: "Who am I?" (in the initial stage) (Pfeiffer & Jones, 1973); identifying fears and conflicts regarding the group (in the transition stage); "Johari Window" (in the working stage) (Luft, 1970); "Coins: Symbolic Feedback" (Pfeiffer & Jones, 1973); and reviewing the group (in the ending stage). The online assignments were completed on the class "Discussion Board" forums of the online course platform Blackboard.² These were based on an exploration of concepts in *The Theory and Practice of Group Psychotherapy* (Yalom & Leszcz, 2005).

Prospective participants were asked, at the onset of the first session, to participate in a research study and to read an informed consent letter that included a description of the procedures and a statement that they may choose to not participate in the data collection while still remaining in the group. The data collection consisted of filling out a four-page packet of questionnaires at the very end of each session, taking 5–10 minutes. Informed consent was given and it was explained that immediately after each session the packets would be distributed and that participants would drop their packets into a box. These procedures were followed at the very end of each session. There were no penalties for declining to participate and no rewards for participating. Volunteers were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" ("Ethical Principles of Psychologists and Code of Conduct," 2002; "2010 Amendments to the 2002 'Ethical principles of psychologists and code of conduct,'" 2010).

Design

Data were collected about the working alliance throughout the group's development. The correlations between the working alliance measure scores for the relationship with the therapist and those for the relationship with the group as a whole were examined. Correlations between ratings of working alliance at all of the eight sessions and outcomes at the end of the group's life were assessed to address the research questions and hypotheses.

Measures

Working Alliance Inventory–Short Revised Form (WAI–SR).—This instrument is a 12-item short form (Hatcher & Gillasp, 2006) of the 36-item Working Alliance Inventory (WAI) originally developed by Horvath and Greenberg (1986). Horvath and Gillasp (2006) reported convergent validity of the WAI–SR with the Helping Alliance Inventory (Horvath & Bedi, 2002) and the California Psychotherapy Alliance Scale (Gaston & Marmar, 1994). It has been shown to have strong internal constancy reliabilities (Cron-

²<http://www.blackboard.com>

bach's α) ranging from .91 to .92 for the total score and .85 to .90 for the subscale scores (Hatcher & Gillaspay, 2006). A recent validation study (Munder, Wilmers, Leonhart, Linster, & Barth, 2010) investigating the psychometric properties of the WAI-SR reported good reliability ($\alpha = .80$) as well as good convergent validity with the Helping Alliance Inventory ($r = .64$).

The WAI-SR was developed to clearly differentiate the "Goal," "Task," and "Bond" dimensions of the working alliance. "Goal" is defined as what the client wishes to gain from therapy, "Task" is defined as the work that the therapist and the client agree needs to be done to achieve his or her goals, and "Bond" is defined as the trust the client has in the therapist to help attain his or her goals. The validation study by Munder, *et al.* (2010) confirmed this factor structure of the WAI-SR.

Respondents in the present study were asked to rate, on a 7-point Likert-type scale, how much they agreed or disagreed with each statement. Statements address one of the three factors that constitute the scale. For example, "We agree on what is important for me to work on" addresses the goal factor; "I believe the way we are working on my problem is correct" addresses the task factor; and "I believe _____ likes me" addresses the "bond" factor. The WAI-SR was adapted slightly in order to meet the needs of this study. Our version asked the participants to rate each of the statements twice; once as the statement pertains to the therapist, and again as it pertains to "the group as a whole." ("Task" statements were rated only once because their wording refers to the work of the participant; for example: "What I am doing in therapy gives me new ways of looking at my problem.")

Outcome measure.—As part of the last-session packet, the participants were asked to rate their agreement on a 7-point Likert-type scale (1: Not at all, 7: Very much) with six statements. These were:

1. Overall, I self-disclosed in this group.
2. Overall, others self-disclosed in this group.
3. Overall, I gave others feedback and support.
4. Overall, others gave me feedback and support.
5. Overall, I felt that I benefited from this group experience.
6. Overall, I felt that others benefited from this group experience.

The six items of this measure had a Cronbach's alpha of .77, which was accepted as satisfactory evidence of reliability. These items were worded so that it was possible to distinguish the participants' perceptions of their own outcomes as well as their perceptions of the outcomes of other members in the group. The items regarding the experiences of others referred

to how each participant believed others experienced the group; i.e., not to any facts that would not be known to the participants.

Demographic questionnaire.—The participants' age ("optional") and sex on the first day's packet were asked. They were also asked to choose a four-digit number that they would write at the top of each subsequent day's packet to maintain their anonymity.

RESULTS

Data Reduction of Outcome Measure

The six items measuring the outcome of group counseling experiences were treated as a single scale. To further validate the psychometric properties of this outcome measure, the factor structure was investigated. A Bartlett's test of sphericity [$\chi^2(15) = 40.5, p < .001$] implied that the correlation matrix of the items was adequate for a factor analysis to be conducted. A principal components analysis was conducted on the six items, using both orthogonal and oblique rotation methods. Due to the relative lack of correlation among factors, orthogonal rotation was selected (Varimax procedure).

In determining the factor structure of the outcome measure, this study considered the Kaiser-Guttman criterion of eigenvalues greater than 1, as well as Cattell's scree plot. In both cases, a two-factor model provided the best solution. With eigenvalues of 2.1 and 1.5, respectively, these two factors accounted for 59.3% of the variance. The first factor (after rotation, constituting 34.3% of the variance) seems to reflect perceived self-improvement. The three items pertained to perceived self-disclosure (Item 1, factor loading of .84), group support (Item 4, .81), and benefitting from the group experience (Item 5, .64). The second factor (after rotation, constituting 25.0% of the variance) focused on the perceived benefits for other group members. The three items for this factor pertained to the perception of how other members benefited from the group experience (Item 6, factor loading of .80), whether other members self-disclosed (Item 2, .67), and whether feedback and support was given to other members (Item 3, .66). Both factors have been previously linked to successful group outcomes. For instance, as pertains to self-improvement, Tschuschke and Dies (1994) found that inpatient therapy group participants who engaged in more self-disclosure had more favorable outcomes, as measured through process ratings and reports of cohesiveness, self-disclosure, feedback, and interpersonal learning-output by both therapists and group members, than those participants who did not self-disclose. In addition, regarding the second factor, it has been found that the more feedback was given and received by group members, the more likely those members were to rate the group experience in a positive manner (Morran, Stockton, & Whittingham, 2004).

Working Alliance with the Therapist and with the “Group as a Whole”

In order to learn if there is a distinction within counseling groups between the members' working alliances with the therapist and with the group as a whole (the first question), correlations were calculated between the members' ratings on the WAI-SR Bond and Goals subscales for the therapist and for the group. The Task subscale is worded so that only the individual respondent's commitment to the task is measured (e.g., “What I am doing in this group gives me new ways of looking at my issues”). The Pearson correlation coefficients of WAI-SR (Bond and Goal subscale) ratings for the therapist with those for the group-as-a-whole were consistently very highly positively correlated ($r > .89$), answering Research Question 1 in the affirmative. Consequently, therapist and group-as-a-whole ratings scores were combined and treated as one variable in further analyses.

Working Alliance Ratings and Counseling Outcomes

To see whether group members' early ratings of the working alliance would be statistically significantly correlated with outcomes at the end of the group's life, correlations between WAI-SR scores (total and subscales) and scores on the outcome measure (total and subscales) were assessed. Table 1 displays the Pearson correlation coefficients of WAI-SR scores at the end of the 1st, 4th, and 8th sessions with the outcome scores gathered at the

TABLE 1
CORRELATIONS (95%CI) OF WORKING ALLIANCE INVENTORY RATINGS AT THREE TIMES
(DAYS 1, 4, 8) WITH OUTCOME MEASURES ON DAY 8

Working Alliance Inventory	Time	M	SD	Pearson r (95%CI)		
				Outcome: Total	Self Outcome	Others' Outcome
Total	Day 1	72.16	13.58	.01	.01	.01
	Day 4	81.83	13.66	.45† (.18, .66)	.50† (.24, .69)	.16
	Day 8	90.07	11.30	.39* (.11, .61)	.53† (.28, .71)	-.02
Bond	Day 1	33.95	5.32	.12	.07	.14
	Day 4	36.17	4.43	.38* (.10, .60)	.42* (.18, .63)	.10
	Day 8	30.87	3.79	.36† (.08, .59)	.49† (.23, .68)	-.04
Goal	Day 1	25.21	7.42	-.06	.01	-.12
	Day 4	30.33	8.02	.44† (.17, .65)	.44† (.17, .65)	.15
	Day 8	34.60	6.68	.43† (.16, .64)	.52† (.27, .70)	.08
Task	Day 1	13.06	3.65	.02	.06	-.04
	Day 4	15.33	3.40	.27	.26	.14
	Day 8	17.57	2.83	-.01	.14	-.28

* $p < .05$ (two-tailed). † $p < .01$ (two-tailed).

end of the last (8th) session. At the end of the 1st session, none of the working alliance scores correlated statistically significantly with the outcome measures (neither the total scores nor the two subscales of one's Self-Outcome or of Others' Outcome). Such correlations were statistically non-significant through sessions 2 and 3. Thus, Hypothesis 1a was not supported.

At the close of the 4th session and thereafter, statistically significant correlations were seen between WAI-SR Total scores and the Total outcome scores ($r_{37} = .45, p < .01$) as well as the Self-Outcome scores ($r_{37} = .50, p < .001$), partly supporting Hypothesis 1b. However, no statistically significant correlations were seen between WAI-SR and Others' Outcome scores. No statistically significant correlations were observed between the WAI-SR Task subscale and any outcomes on any of the days.

At the conclusion of the 8th session, a pattern of correlations was similar to that of session 4. For example, the WAI-SR Total score correlated statistically significantly with Total Outcome scores ($r_{36} = .39, p < .01$) and with the Self-Outcome score ($r_{36} = .53, p < .001$). Once again, WAI-SR correlations with the Others' Outcome subscale were statistically non-significant.

Hypothesis 1a, that early ratings of the working alliance would be correlated with outcomes, was supported for the overall alliance as well as for the Bond and Goal components of the alliance as early as session 4. However, the Task component appeared to be unrelated to outcomes. Hypothesis 1b was that ratings of the working alliance at the end of the last session would be positively correlated with outcomes. This was the case except for the Task factor of the working alliance.

Finally, to assess whether a strong working alliance would be related to self-disclosure in a group (Hypothesis 2), the members' WAI-SR total scores on the last day was correlated with the single outcome measure item of "Overall, I self-disclosed in the group." A statistically significant positive correlation was found between these WAI-SR scores and responses to this single item ($r_{36} = .52, p < .001$).

DISCUSSION

Based on earlier research that showed the importance of the working alliance in therapy, the importance of the working alliance within group therapy was studied. For example, Castonguay, *et al.* (2006) cited several meta-analyses of individual psychotherapy outcomes showing effect sizes of the working alliance ranging from .22 to .26. A recent meta-analysis of 190 independent studies of the relationship between alliance and outcomes showed an aggregate effect size of .275 (Horvath, Del Re, Flückiger, & Symonds, 2011). Therefore, asking some basic questions began an ongoing project focused on whether group members distinguish between their alliances toward the therapist and the group as a whole; how quickly in

the life of a therapy group the working alliance would be related to outcomes; which aspects of the working alliance were most correlated with therapy outcomes; and, finally, the relationship of the working alliance with members' self-disclosure in the group.

Given that the working alliance was being examined within a group context, the fundamental question was whether such alliances tend to be made with the therapist or with the group. The very high correlations (.89 to .95) of ratings of these two aspects of the alliance seem to indicate that there is no distinction between them in the group members' minds. However, this finding may be an artifact of the way the ratings were made on the WAI-SR form. The participants entered their responses to each WAI-SR item for the therapist and for the group-as-a-whole alongside each other. This pairing may have created a bias toward simply entering the same rating twice, without giving the second response much thought. Of course, it is possible that the ratings truly reflect similar judgments and that the two forms of the alliance are indeed very similar to each other. Further inquiry is called for to clarify this finding. In the present study, the therapist and group-as-a-whole ratings scores were combined in the further analyses.

The group members' perceptions of the working alliance were found to be statistically significantly correlated with outcomes. Earlier research had shown that in individual therapy, ratings of the working alliance after the third session were related with outcomes (see the meta-analysis by Horvath & Symonds, 1991). This was found to occur after the fourth session; perhaps therapeutic alliances take a little longer to become established in group therapy. Further study of this issue may be warranted.

Participants' reports of their working alliances correlated statistically significantly with outcomes beginning at session 4 and through to the last session. When the two subscales of the outcome measure were examined, the group members' self-reported working alliance was found to correlate with their perceptions of their own outcomes. This is not the case for their perceptions of other members' outcomes. As might be expected, experiencing a strong working alliance is not related to how one perceives the outcomes of other people in the group.

Due to the fact that the working alliance is composed of three factors (Bond, Goal, Task) the relationships of each factor with outcomes was examined. An interesting pattern emerged: the Bond and Goals aspects of the working alliance both were statistically significantly correlated with outcomes. However, the Task aspect (group members' perceptions that the tasks of the group experience are appropriate) was not related to outcomes. This seems to point to the interesting possibility that the emotional bond and the agreement on goals in group therapy are more important than the particular tasks in a group experience.

The group members' working alliance in relation to their self-disclosure was also studied. When the working alliance ratings were correlated with responses to the single outcome measure item "Overall, I self-disclosed in the group," a significant relationship was found between the strength of the reported working alliance and the degree of reported self-disclosure. Although this finding supports that of Kelly and Yuan (2009), it can only be seen as a preliminary finding since it is based on a single-item measure.

Limitations and Conclusion

Some important limitations should be noted. These include the limited nature of outcome measures utilized. Future research could add more established measures such as the Outcome Questionnaire (Wells, Burlingame, Lambert, Hoag, & Hope, 1996). In addition, the formatting of the WAI-S form, as it was adapted for this study, may not be optimal for distinguishing between the participants' perceptions of the leader and the group as a whole. The ratings of the working alliance with the therapist should be formatted as distinct items from those of the alliance with the group as a whole to get more accurate ratings of these two perceptions. Additional research is needed to establish convergent validity of the outcome measure with established outcome measures as well as reliability and validity of the adapted version of the WAI-SR that was used in this study. Future research with group counseling could also include the therapist's perceptions of both the working alliance and the outcomes.

Another limitation relates to the generalizability of the present findings. The group leader was the teacher in these courses. As a result, such biases as acquiescence and demand characteristics may have influenced the ratings of students. Further studies with different arrangements will be needed to establish the generalizability of the present results.

Overall, the present findings can be seen as supporting the importance of a key aspect of the therapeutic relationship in group therapy, the members' working alliance. The results of the present study support the hypothesis that the strength of the working alliance, as perceived by group members, is strongly related to self-reported outcomes.

REFERENCES

- 2010 AMENDMENTS TO THE 2002 'ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT.'
(2010) *American Psychologist*, 65(5). DOI: 10.1037/a0020168
- ARNOW, B. A., BLASEY, C., MANBER, R., CONSTANTINO, M. J., MARKOWITZ, J. C., KLEIN, D. N., THASE, M. E., KOCIS, J. H., & RUSH, A. J. (2007) Dropouts versus completers among chronically depressed outpatients. *Journal of Affective Disorders*, 97(1-3), 197-202. DOI: 10.1016/j.jad.2006.06.017
- ASAY, T. P., & LAMBERT, M. J. (1999) The empirical case for the common factors in therapy: quantitative findings. In M. A. Hubble, B. L. Duncan, & S. D. Miller (Eds.), *The heart and soul of change: what works in therapy*. Washington, DC: American Psychological Association. Pp. 23-55.

- BAKALI, J. V., BALDWIN, S. A., & LORENTZEN, S. (2009) Modeling group process constructs at three stages in group psychotherapy. *Psychotherapy Research*, 19(3), 332-343. DOI: 10.1080/10503300902894430
- BEUTLER, L. E., MALIK, M., ALIMOHAMED, S., HARWOOD, T. M., TALEBI, H., & NOBLE, S. (2004) Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*. (5th ed.) New York: Wiley. Pp. 227-306.
- BILODEAU, C., SAVARD, R., & LECOMTE, C. (2010) Examining supervisor and supervisee agreement on alliance: is shame a factor? *Canadian Journal of Counselling and Psychotherapy*, 44(3), 272-282.
- BORDIN, E. S. (1979) The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252-260. DOI: 10.1037/h0085885
- CASTONGUAY, L. G., CONSTANTINO, M. J., & HOLTFOORTH, M. G. (2006) The working alliance: where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 271-279. DOI: 10.1037/0033-3204.43.3.271
- CONNORS, G. J., CARROLL, K. M., DICLEMENTE, C. C., LONGABAUGH, R., & DONOVAN, D. M. (1997) The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. *Journal of Consulting and Clinical Psychology*, 65(4), 588-598. DOI: 10.1037/0022-006x.65.4.588
- CONSTANTINO, M. J., SCHWAIGER, E. M., SMITH, J. Z., DEGEORGE, J., MCBRIDE, C., RAVITZ, P., & ZUROFF, D. C. (2010) Patient interpersonal impacts and the early therapeutic alliance in interpersonal therapy for depression. *Psychotherapy: Theory, Research & Practice*, 47(3), 418-424. DOI: 10.1037/a0021169
- COREY, G. (2008) *Theory and practice of group counseling*. (7th ed.) Belmont, CA: Wadsworth/Thomson Learning.
- CROWE, T. P., & GRENYER, B. F. S. (2008) Is therapist alliance or whole group cohesion more influential in group psychotherapy outcomes? *Clinical Psychology & Psychotherapy*, 15(4), 239-246. DOI: 10.1002/cpp.583
- D'IUSO, D., BLAKE, E., FITZPATRICK, M., & DRAPEAU, M. (2009) Cognitive errors, coping patterns, and the therapeutic alliance: a pilot study of in-session process. *Counselling & Psychotherapy Research*, 9(2), 108-114. DOI: 10.1080/14733140902804276
- DUNCAN, B. L. (2010) The heart and soul of change: the biggest bang for your buck. In B. L. Duncan (Ed.), *On becoming a better therapist*. Washington, DC: American Psychological Association. Pp. 123-143.
- ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT. (2002) *American Psychologist*, 57(12), 1060-1073. DOI: 10.1037/0003-066x.57.12.1060
- GASTON, L., & MARMAR, C. R. (1994) The California Psychotherapy Alliance Scales. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: theory, research, and practice*. Oxford, UK: John Wiley & Sons. Pp. 85-108.
- HATCHER, R. L., & GILLASPY, J. A. (2006) Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, 16(1), 12-25. DOI: 10.1080/10503300500352500
- HORVATH, A. O., & BEDI, R. P. (2002) The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: therapist contributions and responsiveness to patients*. New York: Oxford Univer. Press. Pp. 37-69.
- HORVATH, A. O., DEL RE, A. C., FLÜCKIGER, C., & SYMONDS, D. (2011) Alliance in individual psychotherapy. *Psychotherapy*, 48(1), 9-16.

- HORVATH, A. O., & GREENBERG, L. S. (1986) The development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: a research handbook*. New York: Guilford Press. Pp. 529-556.
- HORVATH, A. O., & SYMONDS, B. D. (1991) Relation between working alliance and outcome in psychotherapy: a meta-analysis. *Journal of Counseling Psychology*, 38(2), 139-149. DOI: 10.1037/0022-0167.38.2.139
- KELLER, S. M., ZOELLNER, L. A., & FEENY, N. C. (2010) Understanding factors associated with early therapeutic alliance in PTSD treatment: adherence, childhood sexual abuse history, and social support. *Journal of Consulting and Clinical Psychology*, 78(6), 974-979. DOI: 10.1037/a0020758
- KELLY, A. E., & YUAN, K-H. (2009) Clients' secret keeping and the working alliance in adult outpatient therapy. *Psychotherapy: Theory, Research, Practice, Training*, 46(2), 193-202. DOI: 10.1037/a0016084
- KNERR, M., BARTLE-HARING, S., MCDOWELL, T., ADKINS, K., DELANEY, R. O., GANGAMMA, R., GLEBOVA, T., GRAFSKY, E., & MEYER, K. (2011) The impact of initial factors on therapeutic alliance in individual and couples therapy. *Journal of Marital and Family Therapy*, 37(2), 182-199. DOI: 10.1111/j.1752-0606.2009.00176.x
- LAMBERT, M. J. (1992) Psychotherapy outcome research: implications for integrative and eclectic therapists. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration*. New York: Basic Books. Pp. 94-129.
- LUBORSKY, L., BARBER, J. P., SIQUELAND, L., & JOHNSON, S. (1996) The revised Helping Alliance questionnaire (HAQ-II): psychometric properties. *Journal of Psychotherapy Practice & Research*, 5(3), 260-271.
- LUFT, J. (1970) *Group processes: an introduction to group dynamics*. Bethesda, MD: National Press Books.
- LUSTIG, D. C., STRAUSSER, D. R., WEEMS, G. H., DONNELL, C. M., & SMITH, L. D. (2003) Traumatic brain injury and rehabilitation outcomes: does the working alliance make a difference? *Journal of Applied Rehabilitation Counseling*, 34(4), 30-37.
- MARTIN, D. J., GARSKE, J. P., & DAVIS, M. K. (2000) Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68(3), 438-450. DOI: 10.1037/0022-006x.68.3.438
- MCBRIDE, C., ZUROFF, D. C., RAVITZ, P., KOESTNER, R., MOSKOWITZ, D. S., QUILTY, L., & BAGBY, R. M. (2010) Autonomous and controlled motivation and interpersonal therapy for depression: moderating role of recurrent depression. *British Journal of Clinical Psychology*, 49(4), 529-545. DOI: 10.1348/014466509x479186
- MORRAN, D. K., STOCKTON, R., & WHITTINGHAM, M. H. (2004) Effective leader interventions for counseling and psychotherapy groups. In J. L. DeLucia-Waack, D. A. Gerrity, C. R. Kalodner, & M. T. Riva (Eds.), *Handbook of group counseling and psychotherapy*. Thousand Oaks, CA: Sage. Pp. 91-103.
- MUNDER, T., WILMERS, F., LEONHART, R., LINSTER, H., & BARTH, J. (2010) Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clinical Psychology & Psychotherapy*, 17(3), 231-239.
- O'DELL, T. (2009) *Strength of the working alliance and subsequent development of the goal, task, and bond between supervisor and supervisee using various supervision modalities*. 69, ProQuest Information & Learning, U.S. Retrieved from <http://rlib.pace.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2009-99050-217&site=ehost-live&scope=site>. Available from EBSCOhost psych database.

- PATTERSON, C. L., UHLIN, B., & ANDERSON, T. (2008) Clients' pretreatment counseling expectations as predictors of the working alliance. *Journal of Counseling Psychology*, 55(4), 528-534. DOI: 10.1037/a0013289
- PEREIRA, T. (2010) *The therapeutic alliance and outcome in anorexia nervosa*. 70, ProQuest Information & Learning, USA. Retrieved from <http://rlib.pace.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=2010-99040-508&site=ehost-live&scope=site>. Available from EBSCOhost psych database.
- PEREIRA, T., LOCK, J., & OGGINS, J. (2006) Role of therapeutic alliance in family therapy for adolescent anorexia nervosa. *International Journal of Eating Disorders*, 39(8), 677-684. DOI: 10.1002/eat.20303
- PFEIFFER, J. W., & JONES, J. E. (1973) *A handbook of structured experiences for human relations training*. Vol. 4. Oxford, UK: Univer. Associates.
- TAFT, C. T., MURPHY, C. M., KING, D. W., MUSSER, P. H., & DeDEYN, J. M. (2003) Process and treatment adherence factors in group cognitive-behavioral therapy for partner violent men. *Journal of Consulting and Clinical Psychology*, 71(4), 812-820. DOI: 10.1037/0022-006x.71.4.812
- TRACEY, T. J., & KOKOTOVIC, A. M. (1989) Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1(3), 207-210. DOI: 10.1037/1040-3590.1.3.207
- TRYON, G. S., & KANE, A. S. (1993) Relationship of working alliance to mutual and unilateral termination. *Journal of Counseling Psychology*, 40(1), 33-36. DOI: 10.1037/0022-0167.40.1.33
- TSCHUSCHKE, V., & DIES, R. R. (1994) Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy*, 44(2), 185-208.
- WAMPOLD, B. E. (2001) *The great psychotherapy debate: models, methods, and findings*. Mahwah, NJ: Lawrence Erlbaum.
- WAMPOLD, B. E. (2007) Psychotherapy: the humanistic (and effective) treatment. *American Psychologist*, 62(8), 857-873. DOI: 10.1037/0003-066x.62.8.857
- WEBB, C. A., DeRUBEIS, R. J., AMSTERDAM, J. D., SHELTON, R. C., HOLLON, S. D., & DIMIDJIAN, S. (2011) Two aspects of the therapeutic alliance: differential relations with depressive symptom change. *Journal of Consulting and Clinical Psychology*, 79(3), 279-283. DOI: 10.1037/a0023252
- WELLS, M. G., BURLINGAME, G. M., LAMBERT, M. J., HOAG, M. J., & HOPE, C. A. (1996) Conceptualization and measurement of patient change during psychotherapy: development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training*, 33(2), 275-283. DOI: 10.1037/0033-3204.33.2.275
- YALOM, I. D., & LESZCZ, M. (2005) *The theory and practice of group psychotherapy*. (5th ed.) New York: Basic Books.

ROMANTIC PARTNERS IN A MARKET PERSPECTIVE: EXPECTATIONS ABOUT WHAT ENSURES A HIGHLY DESIRABLE PARTNER¹

FELIPE N. CASTRO, WALLISEN T. HATTORI, MARIA EMÍLIA YAMAMOTO,
AND FÍVIA A. LOPES

*Graduate Program in Psychobiology
Federal University of Rio Grande do Norte*

Summary.—This study used the biological market perspective and influential statistical models from the marketing field to investigate males' and females' expectations regarding which combination of characteristics are most relevant in ensuring desirable partnerships for same-sex individuals. Thus, 358 Brazilian undergraduates assessed eight descriptions of same-gender stimulus targets (formulated with different levels of physical attractiveness, social skills, and current or prospective social status) and evaluated the overall desirability of the targets' expected or probable partners. From the possible combinations, three groups emerged: for one group, mainly composed of men, status characteristics were the most important attributes; for the others, mostly composed of women, social skills or physical characteristics were identified as most important in appealing to a desirable partner. This work expands the understanding of variability in male and female romantic expectations, and its implications are discussed from an evolutionary perspective.

In all human relationships, trustworthiness and cooperativeness are highly valued, and the importance of other characteristics depends on the goals people hope to achieve together (Cottrell, Neuberg, & Li, 2007). Concerning how romantic partners are chosen, physical attractiveness, financial status, and social skills are extremely relevant (Fletcher & Simpson, 2000; Sprecher & Regan, 2002; Lee, Loewenstein, Ariely, Hong, & Young, 2008). In such relationships, people become more selective as investment and commitment increase (Fletcher, Tither, O'Loughlin, Friesen, & Overall, 2004; Woodward & Richards, 2005; Brase, 2006; Castro & Lopes, 2011).

Although the most desirable romantic partner is supposed to be the one with the greatest number of positive attributes, the relevance of each characteristic, and therefore its value in partner choice, varies according to what each trait represents. Sexual strategies theory (Buss & Schmitt, 1993) suggests that both men and women have evolved distinct psychological mechanisms that underlie their mate preferences. Typically, male preference is related to female physical characteristics, and female preferences

¹Address correspondence to Felipe Nalon Castro, Departamento de Fisiologia, Programa de Pós-graduação em Psicobiologia, Universidade Federal do Rio Grande do Norte, Campus Universitário, Caixa Postal 1511, Lagoa Nova, 59078-970 Natal, RN, Brazil or e-mail (castrofn@gmail.com).

are related to male social status and the ability to acquire resources. From an evolutionary perspective, these preferences were selected for because they yielded the best outcomes in terms of reproductive success, given the differential investments in offspring exhibited and the environmental challenges faced (Trivers, 1972). Although the evidence shows some regularity in sex preference patterns, variations are expected and observed in reflection of the individual adjustment to local conditions (Gangestad & Simpson, 2000; Marlowe, 2004; Pillsworth, 2008). Based on this flexibility, one may expect an overlap in the strategies of men and women under certain circumstances.

The process of romantic partner selection shares several characteristics with market systems; for instance, it includes evaluations, contextual influences, and choices. In fact, a biological market model (Noë & Hammerstein, 1995) can easily be applied to romantic relationships, where partnerships are established by the evaluation of specific characteristics of potential partners. In market jargon, traits are treated as commodities; traders (men and women) assume the role of consumers when they evaluate potential partners, and then assume the role of products when they are evaluated by the potential partners. However, these evaluations do not occur randomly. Male and female assessments follow specific strategies based on the asymmetry of parental investment in potential offspring (Trivers, 1972; Buss & Schmitt, 1993; Pawlowski, 2000; Castro, Hattori, & Lopes, 2012). In this sense, the use of a market research approach to romantic relationship research can bring new insights and uncover new information about this widely studied process.

In the present study, the biological market perspective was applied to investigate male and female expectations regarding which among a set of characteristics (physical attractiveness, social skills, and current or prospective social status) is most relevant in ensuring desirable partnerships for individuals of the same sex. The idea behind the hypotheses was that, although men and women should expect that different attributes are the most relevant for their own sex, there would be some overlap of expectations; moreover, the way people perceive themselves ought to affect their expectations. This work explored the paradigm of the mating market using two of the most influential statistical models used in the marketing field: conjoint analysis and cluster analysis. In comparison to directly contrasting previously defined groups, this approach allows new groups to emerge, thereby exploring people's beliefs and expectations in the mating market in a more informative manner.

Hypothesis 1. Participants of both sexes can be grouped because they share the belief that a specific set of attributes is the most relevant in ensuring a desirable partner.

Hypothesis 2. Most male participants are expected to attribute more desirable partnerships to those males enjoying higher financial status, and most female participants are expected to attribute more desirable partnerships to those females enjoying higher physical attractiveness.

Hypothesis 3. Participants will identify as most relevant the same characteristics on which they rate themselves most highly.

METHOD

Participants

A total of 358 Brazilian university students, ages 18 to 29 years, took part in the study. The sample was composed of 132 men (M age = 21.9 yr., SD = 2.7) and 226 women (M age = 21.2 yr., SD = 2.6). According to the Brazilian skin color classification (Instituto Brasileiro de Geografia e Estatística, 2010), 45.2% of these participants were white, 38.9% were dark-skinned (mixed-race or pardo), 8.4% were yellow (Asian individuals: Japanese, Chinese, and Korean), 5.3% were black, and 2.2% were indigenous. In the sample, 90.2% of the participants reported being attracted to the opposite sex, 4.8% to the same sex, and 5.0% to both sexes. Data collection was performed at one public and two private universities in Natal, Brazil. In this study, all undergraduate students were invited to participate during their academic activities. For those who decided to participate, participation was voluntary, and no academic or financial incentive was offered in exchange.

Procedure

The study was conducted in accordance with ethical guidelines, under the approval of the institutions involved, and with the participants' consent. All items included in the questionnaires (i.e., protocols, descriptions, and explanations) were administered in Portuguese. Participants responded to an anonymous, individual questionnaire containing the descriptions of eight stimulus targets of the same gender as themselves. They were also asked to rate the targets' profiles (to ensure the validity of the measures), to provide a self-perception rating, and to answer personal information questions.

Formulation and Presentation of Target Descriptions

For the purposes of the study, the eight descriptions of target individuals were presented in paragraphs and were developed to cover all possible combinations of three categories: physical attractiveness, social skills, and social status, each presented in one of two levels (High or Low), in accordance with a full-profile (full-concept) approach ($2 \times 2 \times 2$). These

attributes were chosen because they relate to traits considered important in the decision-making process of selecting partners, which, according to several studies, fall into three main categories: attractiveness and vitality, warmth and trustworthiness, and status and resources (Gangestad & Simpson, 2000; Fletcher, *et al.*, 2004). To mitigate effects of order, the descriptions were presented to the participants in a random order. To ensure the validity of the descriptions, participants also rated the profiles on nine traits using a 10-point Likert-type scale (0: Very poor trait expression to 9: Very good trait expression): attractiveness of face, attractiveness of body, health (traits related to physical attractiveness); sociability, agreeableness, sincerity (traits related to social skills); and good financial status, ambition/hard worker, and intelligence (traits related to current/prospective social status). Descriptions were formulated with high or low attributes for the three sets of features for male and female participants, as shown in the Appendix.

Assessment of Desirability of Targets

In the questionnaire, the participants were informed that each target had a romantic partner and were asked to rate the overall desirability of the target's partner. This indirect evaluation (the evaluation of quality of a target associated with the quality of their partner) was used to avoid any gender bias that might occur by direct assessment of targets of the same sex as the participants. To evaluate the overall desirability expected for the targets' partners, the participants answered a semantic differential scale in which they were asked to mark a reference point in a 100 mm straight line, ranging from (left) "Not desirable at all" to (right) "Extremely desirable." In this scale, lower scores represented low evaluations of the target's partner.

Self-perception Protocol

In the self-perception protocol, participants were asked to describe themselves on the following nine traits using a 10-point Likert-type scale (0: Very poor trait expression to 9: Very good trait expression): attractiveness of face, attractiveness of body, health, sociability, agreeableness, sincerity, good financial status, ambition/hard worker and intelligence. The first three characteristics referred to physical attractiveness, the next three to social skills, and the last to status. (For an example of a similar self-perception protocol with a demographically equivalent sample, see Castro, *et al.*, 2012.) At the beginning of the procedure, participants were informed that a higher point allocation to a given trait would indicate a higher score on the 10-point Likert-type scale and that, on the 100 mm visual analog scale, the lower the score, the lower the overall desirability.

Analysis

General linear model (GLM) tests were performed to verify if the eight targets were perceived by the participants according to the intent of

the dimensions used in their formulation. In these tests, the evaluations of each target's characteristics acted as the dependent variable, while the target's description (form SD1 to SD8) was the fixed factor.

To assess the relative importance placed on the same-sex target profiles' physical attractiveness, social skills and social status in ensuring high quality partners, a full-profile (full-concept) conjoint analysis was performed for each participant. This method is a favored methodology in marketing and is used to investigate buyers' choices among competing products (Green, Krieger, & Wind, 2001). It is applied specifically to understand how respondents assess the value of certain types of objects (e.g., ideas, products, or services), decomposing the overall assessment of an object into the values assigned to the attributes that compose each object. In this analysis, a factorial design ($2 \times 2 \times 2$) based on the eight profiles and the overall desirability expected of the eight targets' partners were used. Each data point was a preference point assigned to the profiles, and the linear model was applied for all factors.

Based on the percentages of each attribute's relevance obtained from the conjoint analysis, the participants' data were assessed using hierarchical cluster analysis to identify the number of mutually exclusive groups that could be observed in the sample, considering only the similarities or differences between them. In this procedure, the number of clusters was identified using the between-groups linkage method with the measure of Euclidean distance squared. The number of clusters was obtained by observing the greatest distance between the clustering coefficients in the agglomeration schedule. Then, participants were classified using K-means cluster analysis, a technique used to classify the sample of individuals into a number of mutually exclusive groups.

Frequency differences of men and women in the clusters were investigated using a chi-square test, and differences in self-perception ratings were investigated using GLM tests for each trait in the participants' self-perception data. Under the GLM, the evaluations of each characteristic acted as dependent variables, while the participant's sex and cluster membership were fixed factors. For all tests, alpha was set at .05.

RESULTS

Taken together, the results from the validation of the profiles supported the hypothesis that the assessed profiles represented the dimensions proposed by the study. The descriptions with high physical attractiveness were described as having more attractive faces and bodies than those with low physical attractiveness; those described as having high social skills were rated as more sociable, agreeable and sincere compared with those described as having low social skills; and targets described as having high social status were rated as having better financial status and higher intelligence compared with those of low social status (Table 1).

TABLE 1
DESCRIPTIVE STATISTICS AND GLM RESULTS WITH BONFERRONI *POST HOC* CONTRASTS OF
CHARACTERISTICS OF TARGETS' DESCRIPTIONS

	Target Description (TD)							
	TD1	TD2	TD3	TD4	TD5	TD6	TD7	TD8
Design								
Physical attractiveness	High	High	High	High	Low	Low	Low	Low
Social skills	High	High	Low	Low	High	High	Low	Low
Current or prospective social status	High	Low	High	Low	High	Low	High	Low
Results								
Physical attractiveness								
Attractive face								
<i>M</i>	7.24 ^A	7.46 ^A	7.29 ^A	5.96 ^B	3.08 ^C	2.63 ^D	1.99 ^E	2.02 ^E
<i>SD</i>	1.83	1.83	1.79	2.31	1.78	1.51	1.50	1.56
Attractive body								857.84*
<i>M</i>	7.07	7.63	6.92	7.75	2.74	2.64	1.77	1.77
<i>SD</i>	1.86 ^B	1.76 ^A	1.94 ^B	1.67 ^A	1.80 ^C	1.58 ^C	1.47 ^D	1.42 ^D
Good health								1,139.77*
<i>M</i>	6.64 ^B	6.22 ^C	5.75 ^D	7.41 ^A	6.27 ^C	5.22 ^E	4.00 ^F	3.73 ^F
<i>SD</i>	1.95	2.03	2.15	1.87	2.19	2.02	2.08	1.93
Social skills								211.44*
Sociability								
<i>M</i>	7.75 ^A	7.29 ^B	3.67 ^C	2.10 ^D	7.47 ^B	7.13 ^B	2.00 ^D	1.98 ^D
<i>SD</i>	1.53	1.78	2.05	1.43	1.76	1.86	1.66	1.72
Agreeableness								1028.33*
<i>M</i>	6.84 ^A	6.95 ^A	2.83 ^B	1.99 ^C	6.78 ^A	6.95 ^A	1.85 ^C	1.85 ^C
<i>SD</i>	1.91	1.93	1.73	1.38	2.07	1.94	1.57	1.71
Sincerity								887.75*
<i>M</i>	6.01 ^A	6.25 ^A	4.12 ^B	3.46 ^C	6.24 ^A	6.28 ^A	2.53 ^D	2.36 ^D
<i>SD</i>	2.25	2.17	2.51	2.47	2.15	2.17	2.29	2.10
Current or prospective social status								261.59*
Good financial status								
<i>M</i>	8.03 ^A	2.58 ^D	8.03 ^A	3.13 ^C	7.52 ^B	2.82 ^D	7.43 ^B	2.15 ^E
<i>SD</i>	1.45	1.40	1.47	1.52	1.53	1.52	1.60	1.39

(continued on next page)

Note.—In rows, similar letters indicate similar means by Bonferroni test ($p < .05$). Boldface font indicates significant differences in *post hoc* analysis between levels (high or low). For all GLM tests $df = 7, 2499$. * $p < .001$.

TABLE 1 (CONT'D)
 DESCRIPTIVE STATISTICS AND GLM RESULTS WITH BONFERRONI *POST HOC* CONTRASTS OF
 CHARACTERISTICS OF TARGETS' DESCRIPTIONS

	Target Description (TD)							
	TD1	TD2	TD3	TD4	TD5	TD6	TD7	TD8
Ambition/Hard worker								
<i>M</i>	7.47 ^A	6.89 ^B	7.06 ^B	4.75 ^D	7.60 ^A	6.73 ^B	5.92 ^C	4.55 ^D
<i>SD</i>	1.80	2.01	2.00	2.32	1.76	1.97	2.16	2.29
Intelligence								
<i>M</i>	7.73 ^A	4.09 ^E	7.35 ^B	3.26 ^F	7.07 ^B	4.59 ^D	5.74 ^C	2.68 ^G
<i>SD</i>	1.56	1.93	1.83	1.88	1.78	1.95	2.12	1.88

Note.—In rows, similar letters indicate similar means by Bonferroni test ($p < .05$). Boldface font indicates significant differences in *post hoc* analysis between levels (high or low). For all GLM tests $df = 7, 2499$. * $p < .001$.

Among a large combination of possible numbers of groups (from each participant individually distributed up to all participants in a single group), the optimal solution of three clusters was observed: Cluster 1, composed of participants who assessed social skills as more important than the other traits in predicting a highly desirable partner for individuals of their own sex; Cluster 2, composed of participants who assessed physical attractiveness as more important than any other characteristic; and Cluster 3, containing participants who assessed social status-related traits as the most important. Differences among the clusters for the mean relevance of physical attractiveness ($F_{2,355} = 332.82, p < .001$), social skills ($F_{2,355} = 358.59, p < .001$) and social status ($F_{2,355} = 180.26, p < .001$) in determining partner desirability were statistically significant (Fig. 1).

Although both sexes were present in all clusters (Table 2), male and female participants were not equally distributed within clusters. The residual analysis revealed that Cluster 1 and Cluster 2 included more women than expected (Cluster 1 showed a higher proportion of women), while Cluster 3 included more men than expected.

In general, the self-perception data indicated that participants of different sexes and in different clusters described themselves in a similar manner (Table 3). Differences between sexes and between clusters were observed only for Ambition/Hard worker ($F_{1,352} = 9.42, p = .002$, partial $\eta^2 = 0.03$; $F_{2,352} = 3.79, p = .02$, partial $\eta^2 = 0.02$, respectively); women ($M = 6.96, SD = 1.87$) had a slightly higher mean self-rating on ambition than men ($M = 6.36, SD = 1.88$), but no significant differences between clusters were observed on *post hoc* analysis (Table 3). No other statistically significant main effects were observed (all $F_{1,352} < 2.87, ps < .09$). Although an interaction effect between sex and cluster was observed for financial status ($F_{2,352} = 3.17, p =$

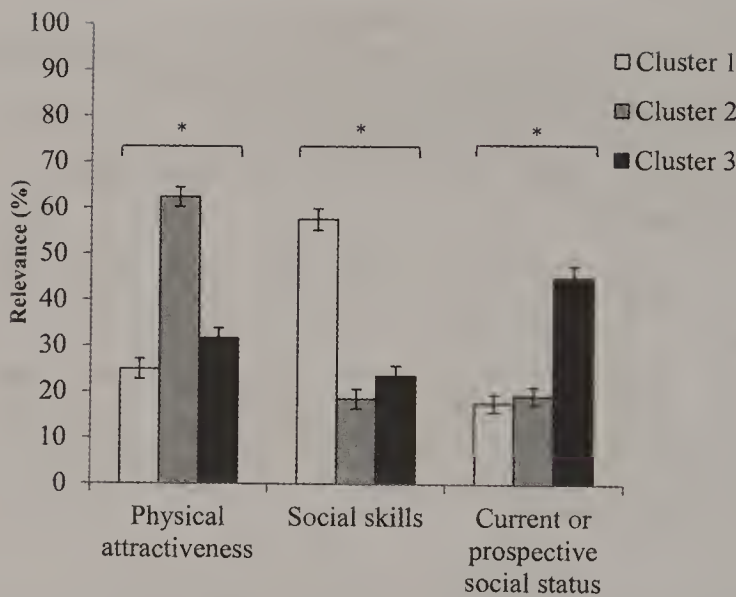


FIG. 1. Mean percentage relevance and error bars (95% confidence intervals) of the traits necessary to ensure a highly desirable partner for same-sex individuals, described by clusters. *Indicates significant differences between Clusters ($p < .05$).

.04, partial $\eta^2 = 0.02$), *post hoc* analysis performed for this characteristic's ratings showed no significant differences. Significant interaction effects were not observed for any other traits (all $F_{1,352} < 2.52$, $ps > .08$).

DISCUSSION

Market and business researchers usually try to understand how the minds of consumers work, especially when consumers have to choose which product they will buy among several similar, competing products (Miller, 2009). People evaluate, select, and choose objects based on their

TABLE 2
CHI-SQUARE INDEPENDENCE TEST RESULTS FOR SEX AND CLUSTERS

Gender		Cluster 1	Cluster 2	Cluster 3	Total
Women	Count	84.00	81.00	61.00	226
	Expected count	77.65	77.02	71.34	226
	Residual	6.35	3.98	-10.34	
Men	Count	39.00	41.00	52.00	132
	Expected count	45.35	44.98	41.66	132
	Residual	-6.35	-3.98	10.34	
Total		123	122	113	358

Note.— $\chi^2(2) = 6.02$, $p = .049$. Boldface font indicates higher proportion of a specific gender in the clusters.

TABLE 3
DESCRIPTIVE STATISTICS OF RESPONDENTS' SELF-DESCRIPTIONS BY CLUSTER

	Attractive- ness		Health	Soci- ability	Agree- ableness	Sin- cerity	Good Financial Status	Ambi- tion/Hard Worker	Intelli- gence
	Face	Body							
Cluster 1									
Female									
<i>M</i>	5.51	5.17	6.37	6.69	7.02	7.27	4.63	6.86	6.37
<i>SD</i>	1.65	1.83	1.86	1.91	1.71	1.74	1.62	1.86	1.60
Male									
<i>M</i>	5.00	4.74	6.69	6.00	6.38	6.54	4.08	5.62	6.05
<i>SD</i>	1.79	1.67	1.66	2.22	2.35	1.83	1.51	2.10	1.65
Cluster 2									
Female									
<i>M</i>	5.64	4.59	6.05	6.63	6.74	7.02	4.38	7.09	6.44
<i>SD</i>	1.76	1.48	1.65	1.62	1.66	1.56	1.50	1.67	1.53
Male									
<i>M</i>	5.37	4.80	6.54	7.07	7.17	6.71	4.54	6.66	6.24
<i>SD</i>	1.55	1.69	1.78	1.62	1.82	1.69	1.36	1.64	1.41
Cluster 3									
Female									
<i>M</i>	5.26	4.93	6.49	6.41	6.61	7.15	3.85	6.92	6.16
<i>SD</i>	2.12	1.90	1.83	2.05	2.05	1.75	1.96	2.15	1.86
Male									
<i>M</i>	5.56	4.77	6.31	6.46	6.67	7.25	4.40	6.69	6.42
<i>SD</i>	1.55	1.63	1.94	1.90	1.85	1.62	1.75	1.77	1.56

needs, their lifestyles and the objects' features. A similar process occurs when people seek romantic relationships. However, in this case, one class of 'traders' (men or women) must assess what the individuals of the other class are seeking in a romantic partner. In addition, they must be able to assess what their potential competitors are offering as romantic partners.

Using the biological market paradigm and statistical techniques widely used in the marketing field, this study was developed to measure participants' judgments of romantic characteristics that an individual should possess to ensure a desirable partner. As predicted, ratings from men and women formed clusters describing different combinations of attributes as being most relevant in ensuring a desirable partner for individuals of their own sex. This result suggests that there is an overlap in male and female strategies when looking for a romantic partner, indicating that the assump-

tion that men and women are expected to have different romantic preferences may not be completely valid. Variability of related beliefs among men and women is usually neglected in methodologies that use the direct contrast between the genders.

Although several individuals from both sexes were observed to share the same beliefs, men and women were not equally distributed among the clusters. Corroborating the sex-specific mate-preference patterns observed in the literature (Buss & Barnes, 1986; Buss, 1989; Buunk, Dijkstra, Fetchenhauer, & Kenrick, 2002; Li, Bailey, Kenrick, & Linsenmeier, 2002; Castro & Lopes, 2011), the results showed that male and female participants were aware of what type of characteristics are the most important for the other sex (more men expected status and more women expected physical attractiveness to be their main assets). Similarly, Pawlowski and Dunbar (1999) found evidence that 'female market value' is related to women's fecundity and that 'male market value' is related to men's earning potential, characteristics associated with physical attractiveness and social status, respectively. Schmitt and Buss (1996) observed that men and women are aware of the characteristics desired by the opposite sex and that they use this information to promote themselves and derogate their competitors in the romantic marketplace.

Investigating self-presentation accuracy in online dating profiles, Toma, Hancock, and Ellison (2008) found that men and women appeared to lie only about some characteristics related to their own sex's attractiveness: women lied more about their weight and men lied more about their height. Checking the participants' self-ratings, the authors found that these inaccuracies were intentional rather than self-deceptive. Similar to the results observed in the present study, their results suggested that men and women have specific beliefs about the attributes that the opposite sex values in romantic dates. Studies that focus on body shape have shown conflicting results: for instance, Fallon and Rozin (1985) observed that men and women err in estimating what the opposite sex would find attractive (*viz*, men overestimate females' preferences for heavier male stature and women overestimate males' preferences for thin women), whereas Bergstrom, Neighbors and Lewis (2004) found that men are accurate in their perceptions of what women find attractive in men but (agreeing with the former work) women believe that men want women to be thinner than men actually report.

The results provided by this work also indicated that social skills were expected by a group of people as the most relevant attributes for individuals of their own sex. Social skills-related characteristics, such as trustworthiness and cooperativeness, seem to be key attributes for almost all human relationships (Cottrell, *et al.* 2007). In fact, there is evidence

that characteristics such as kindness or understanding can be the most important characteristics sought in a mate (Buss, 1985). In the given context, these traits are important because they signal investment in the relationship, which is desired and expected in romantic relationships of long duration (Fletcher, *et al.*, 2004).

Contradicting the last above-mentioned prediction, people with different expectations showed no substantial differences in their self-perceptions. This result suggests variables other than the manner in which people describe themselves could be more relevant in 'mating market' expectations. As previously shown, gender seems to be an important variable influencing certain expectations. Although cultural values and local environmental conditions were not explored in the present study, these variables could also be modulating expectations related to mating behavior. It is important to emphasize that a well-calibrated perception of what the opposite sex is looking for in romantic relationships can help individuals invest and display the most advantageous attributes to potential partners. One could also speculate that misperceptions of what the opposite sex values could lead to the promotion of the wrong characteristics in self-presentation, which in turn could decrease the likelihood of attracting a romantic partner.

Limitations of the current study include the nature of the sample, comprising university undergraduates in their early 20s who are mostly financially dependent on their parents. The use of samples from outside the university, with different financial backgrounds and of different age groups, would certainly improve the generalizability of the results and increase variability in self-perception. With respect to the methodology, new studies using different targets and the use of other specific characteristics in the target descriptions could also reveal important information about people's expectations. More objective self-perception protocols (mate-value questionnaires) or even personality and self-esteem measures could also reveal more detailed effects and help describe why the groups have different expectations and why some men and women could be grouped on seemingly similar beliefs.

Notwithstanding these limitations, this study advances the understanding of mate-selection processes by demonstrating that people believe distinct characteristics help individuals of their own sex attract a desirable partner and that groups of men and women share similar expectations. It added evidence regarding the influence of gender on these expectations, although no evidence supporting the effect of self-perception was observed. This study also indicates that, for a certain group of people, social characteristics may be most relevant to attracting a romantic partner. These results are important because they contribute to the understanding of peo-

ple's mating expectations, which can have direct implications on mate competition, investment, and the characteristics people seek to display to increase their own mate value. Finally, this study also aimed to extend the applications of conjoint and cluster methodologies to human 'mating market' inferences and the underlying drivers of human mating behavior.

REFERENCES

- BERGSTROM, R. L., NEIGHBORS, C., & LEWIS, M. A. (2004) Do men find "bony" women attractive? Consequences of misperceiving opposite sex perceptions of attractive body image. *Body Image*, 1, 183-191. DOI: 10.1016/S1740-1445(03)00025-1
- BRASE, G. L. (2006) Cues of parental investment as a factor in attractiveness. *Evolution and Human Behavior*, 27, 145-157. DOI: 10.1016/j.evolhumbehav.2005.06.003
- BUSS, D. M. (1985) Human mate selection. *American Scientist*, 73, 47-51.
- BUSS, D. M. (1989) Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, 12, 1-49. DOI: 10.1017/S0140525X00023992
- BUSS, D. M., & BARNES, M. (1986) Preferences in human mate selection. *Journal of Personality and Social Psychology*, 50(3), 559-570. DOI: 10.1037/0022-3514.50.3.559
- BUSS, D. M., & SCHMITT, D. P. (1993) Sexual strategies theory: an evolutionary perspective on human mating. *Psychological Review*, 100, 204-232. DOI: 10.1037/0033-295X.100.2.204
- BUUNK, B. P., DIJKSTRA, P., FETCHENHAUER, D., & KENRICK, D. T. (2002) Age and gender differences in mate selection criteria for various involvement levels. *Personal Relationship*, 9, 271-278. DOI: 10.1111/1475-6811.00018
- CASTRO, F. N., & LOPES, F. A. (2011) Romantic preferences in Brazilian undergraduate students: from the short term to the long term. *Journal of Sex Research*, 48(5), 479-485. DOI: 10.1080/00224499.2010.506680
- CASTRO, F. N., HATTORI, W. T., & LOPES, F. A. (2012) Relationship maintenance or preference satisfaction? Male and female strategies in romantic partner choice. *Journal of Social, Evolutionary & Cultural Psychology*, 6, 217-226.
- COTTRELL, C. A., NEUBERG, S. L., & LI, N. P. (2007) What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology*, 92, 208-231. DOI: 10.1037/0022-3514.92.2.208
- FALLON, A. E., & ROZIN, P. (1985) Sex differences in perceptions of desirable body shape. *Journal of Abnormal Psychology*, 94(1), 102-105.
- FLETCHER, G. J., & SIMPSON, J. A. (2000) Ideal standards in close relationships: their structure and functions. *Current Directions in Psychological Science*, 9, 102-105. DOI: 10.1111/1467-8721.00070
- FLETCHER, G. J., TITHER, J. M., O'LOUGHLIN, C., FRIESEN, M., & OVERALL, N. (2004) Warm and homely or cold and beautiful? Sex differences in trading off traits in mate selection. *Personality and Social Psychology Bulletin*, 30, 659-672. DOI: 10.1177/0146167203262847
- GANGESTAD, S. W., & SIMPSON, J. A. (2000) The evolution of human mating: trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, 23, 573-644. DOI: 10.1017/S0140525X0000337X

- GREEN, P. E., KRIEGER, A. M., & WIND, Y. (2001) Thirty years of conjoint analysis: reflections and prospects. *Interfaces*, 31(3), 55-73. DOI: 10.1287/inte.31.3s.56.9676
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. (2010) Síntese de indicadores sociais: uma análise das condições de vida da população Brasileira. *Estudos & Pesquisas: Informação Demográfica e Socioeconômica* [Synthesis of social indicators: an analysis of the living conditions of the Brazilian population. *Studies & Surveys: Demographic and Socioeconomic Information*]. Rio de Janeiro, Brazil: Author. (Número 27).
- LEE, L., LOEWENSTEIN, G., ARIELY, D., HONG, J., & YOUNG, J. (2008) If I'm not hot, are you hot or not? Physical-attractiveness evaluations and dating preferences as a function of one's own attractiveness. *Psychological Science*, 19(7), 669-677. DOI: 10.1111/j.1467-9280.2008.02141.x
- LI, N. P., BAILEY, J. M., KENRICK, D. T., & LINSSENMEIER, J. A. (2002) The necessities and luxuries of mate preferences: testing the tradeoffs. *Journal of Personality and Social Psychology*, 82(6), 947-955. DOI: 10.1037/0022-3514.82.6.947
- MARLOWE, F. W. (2004) Mate preferences among Hadza hunter-gatherers. *Human Nature*, 15, 365-373. DOI: 10.1007/s12110-004-1014-8
- MILLER, G. (2009) *Spent: sex, evolution, and consumer behavior*. New York: Viking.
- NOË, R., & HAMMERSTEIN, P. (1995) Biological markets. *Trends in Ecology & Evolution*, 10(8), 336-339. DOI: 10.1016/S0169-5347(00)89123-5
- PAWLOWSKI, B. (2000) The biological meaning of preferences on the human mate market. *Anthropological Review*, 63, 39-72.
- PAWLOWSKI, B., & DUNBAR, R. I. M. (1999) Impact of market value on human mate choice decisions. *Proceedings of the Royal Society of London*, 266, 281-285. DOI: 10.1098/rspb.1999.0634
- PILLSWORTH, E. G. (2008) Mate preferences among the Shuar of Ecuador: trait rankings and peer evaluations. *Evolution and Human Behavior*, 29(4), 256-267. DOI: 10.1016/j.evolhumbehav.2008.01.005
- SCHMITT, D. P., & BUSS, D. M. (1996) Strategic self-promotion and competitor derogation: sex and context effects on the perceived effectiveness of mate attraction tactics. *Journal of Personality and Social Psychology*, 70(6), 1185-1204. DOI: 10.1037/0022-3514.70.6.1185
- SPRECHER, S., & REGAN, P. C. (2002) Liking some things (in some people) more than others: partner preferences in romantic relationships and friendships. *Journal of Social and Personal Relationships*, 19, 463-481. DOI: 10.1177/0265407502019004048
- TOMA, C. L., HANCOCK, J., & ELLISON, N. (2008) Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin*, 34(8), 1023-1036. DOI: 10.1177/0146167208318067
- TRIVERS, R. L. (1972) Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man, 1871-1971*. Chicago, IL: Aldine. Pp. 136-179.
- WOODWARD, K., & RICHARDS, M. H. (2005) The parental investment model and minimum mate choice criteria in humans. *Behavioral Ecology*, 16(1), 57-61. DOI: 10.1093/beheco/arh121

APPENDIX

The design included eight descriptions (TD) of targets on three features: physical attractiveness, social skills, and social status:

Feature	Target Description (TD)							
	TD1	TD2	TD3	TD4	TD5	TD6	TD7	TD8
Physical attractiveness	High	High	High	High	Low	Low	Low	Low
Social skills	High	High	Low	Low	High	High	Low	Low
Social status	High	Low	High	Low	High	Low	High	Low

Descriptions for each feature for male respondents were as follows:

Feature	Low
Physical attractiveness	Women consider M. H. terrible looking.
Social skills	He is self-centered and unreliable. He does not have many friends, and even his closest friends do not enjoy spending time with him.
Social status	He works as a waiter in a café that is near his home. Although he enjoys his job, he does not earn much money.
High	
Physical attractiveness	Given his physical characteristics, many people find that D. F. is a very attractive man; some say it's hard to take their eyes off him.
Social skills	D. F. is an easy-going, competent man. He is always helpful to other people even if he has to stop what he is doing.
Social status	He became a government diplomat a few years ago. Now he has a good salary and few expenses, so he has a lot of money to spend however he likes.

For female participants, the gender of the targets was changed.

SOCIAL SUPPORT, WORK-FAMILY CONFLICT, AND EMOTIONAL EXHAUSTION IN SOUTH KOREA¹

SOOJIN LEE AND SECKYOUNG LORETTA KIM

*College of Business Administration
Seoul National University*

EUN KYUNG PARK

CyberMBA Co., Ltd., Korea

SEOKHWA YUN

*College of Business Administration
Seoul National University*

Summary.—With an increase of female workforce and dual-earner families, work-family conflict has received particular attention. Using a sample of 159 employees in South Korea, this study examined whether work-family conflict mediated the relationship between social support and emotional exhaustion. Supervisor and family support were found to be related negatively to two different aspects of work-family conflict, i.e., work interference with family and family interference with work, respectively. Also, each dimension of work-family conflict was associated with employees' emotional exhaustion. The relationship between supervisor support and emotional exhaustion was mediated by work interference with family; whereas, the relationship between family support and emotional exhaustion was mediated by family interference with work. Implications and future research directions are discussed.

Recently, research in work and family issues has 'exploded,' due to the increased number of female workers in the labor force and the prevalence of dual-earner families (Amstad, Meier, Fasel, Elfering, & Semmer, 2011; Odle-Dusseau, Britt, & Greene-Shortridge, 2012). Scholars have paid increasing attention to work-family conflict, particularly its negative consequences (Amstad, *et al.*, 2011). For example, employees with high work-family conflict are likely to experience work stress such as emotional exhaustion. Since emotional exhaustion has negative effects on the individual as well as the organization, it is important to investigate when and why employees experience emotional exhaustion (Cropanzano, Rupp, & Byrne, 2003). Employees who receive low social support are likely to feel work-family conflict, which in turn, may lead to emotional exhaustion. Social support and work-family conflict can be separated into work and non-work domains, following an extant research area (Byron, 2005; Amstad, *et al.*, 2011).

Work-family conflict refers to "a form of inter-role conflict in which the role pressures from the work and family domains are mutually incompatible in some respect" (Greenhaus & Beutell, 1985, p. 77). Many stud-

¹Address correspondence to Seokhwa Yun, College of Business Administration, Seoul National University, 1 Gwanakno, Gwanak-gu, Seoul 151-916, Korea (ROK) or e-mail (syun@snu.ac.kr).

ies have suggested that work-family conflict can be categorized into work interference with family and family interference with work (Gutek, Searle, & Klepa, 1991; Adams, King, & King, 1996; Frone, 2003; Livingston & Judge, 2008; Yu, Lee, & Tsai, 2010). Although prior research has suggested these concepts are correlated (Byron, 2005; Mesmer-Magnus & Viswesvaran, 2005), Mesmer-Magnus and Viswesvaran (2005) concluded that the two discrete measures have enough unique variance to warrant independent examinations despite some overlap. Following their suggestion, this study focuses on both work interference with family and family interference with work domains in investigating the effect of work-family conflict.

Organizational research has documented social support in the work setting (Van Yperen & Hagedoorn, 2003; Halbesleben, 2006; Bacharach, Bamberger, & Biron, 2010;) and, in particular, its buffering effect on work-family conflict (Michel, Mitchelson, Pichler, & Cullen, 2010, Michel, Kotrba, Mitchelson, Clark, & Baltes, 2011), although, obviously, employees also receive support from the non-work domain (Carlson & Perrewé, 1999). At the workplace, supervisors may have powerful influences on the way employees better integrate work and family roles (Nasurdin & Hsia, 2008). Supervisors should seek balance between employees' work and family matters (Ballout, 2008). Support from family can help to reduce work-family conflict in individuals who juggle work and family roles (Carlson & Perrewé, 1999). Thus, this study focuses on support from both supervisor and family as important resources for employees to achieve work-family balance in the work and non-work domains (Lapierre & Allen, 2006; Michel, *et al.*, 2011).

Emotional exhaustion may lead to negative consequences for individuals as well as organizations (Maslach & Jackson, 1981; Boles, Johnston, & Hair, 1997; Cropanzano, *et al.*, 2003). Emotional exhaustion refers to a feeling of being emotionally depleted and exhausted by the various reasons relevant to the work (Maslach & Jackson, 1981). This study focuses on emotional exhaustion as an outcome variable, because it is widely accepted as the "core" of job burnout (Maslach, 1982; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001; Shirom, 2003). Moreover, recent studies have demonstrated that emotional exhaustion is related to many forms of poor attitude and reduced job performance (Cropanzano, *et al.*, 2003). Based on the above discussion, this study examines whether two different aspects of work-family conflict mediate the relations between work and non-work social support and emotional exhaustion.

According to the matching hypothesis (Amstad, *et al.*, 2011), work-related antecedents tend to have a stronger effect on work interference with family than family interference with work. For example, the more time individuals spend at work, the more likely it is that their work may inter-

fere with their family life. The proposed mediating relations are grounded on Hobfoll's (1989) conservation of resources theory, which posits that individuals seek to gain and keep resources and, when they perceive excessive demands or insufficient resources, stressful consequences such as work-family conflict may result. The theory suggests that interference between employees' work and home lives depletes their resources for problem solving and coping. Anything that acts to replenish resources should result in lowered work-family conflict. Grandey and Cropanzano (1999) suggested that those who have more resources, such as support, will suffer less stress and reduced work-family conflict.

Domain-specific resources may reduce occurrence of employees' domain-specific conflict. For example, Michel and his colleagues (2011) found that work social support is associated with work interference with family, while family social support is an antecedent of family interference with work. Accordingly, supervisor support could be an effective resource for employees to manage work interference with family by providing adequate resources, which in turn may lead to lower emotional exhaustion.

Hypothesis 1. Work interference with family will mediate the relationship between supervisor support and emotional exhaustion.

In line with above reasoning about domain specificity, the family-related antecedents may affect family interference with work more than work interference with family. Many researchers have proposed that social support from the non-work domain, such as a spouse or other family members, is important in lowering family interference with work (Adams, *et al.*, 1996; Bernas & Major, 2000). For example, Demerouti, Geurts, and Kompier (2004) found that individuals who have low social support and control in their home situations reported high family interference with work. Family support makes employees' situations less stressful by supplying emotional aid, instrumental assistance, or possibly providing considerable adaptability or control over their situations, reducing emotional exhaustion.

Hypothesis 2. Family interference with work will mediate the relationship between family support and emotional exhaustion.

METHOD

Participants and Procedure

The researchers met all the participants personally to describe the procedure for completing the survey and to inform them of the purpose of this study. All respondents were assured that complete confidentiality was guaranteed. Also, they were asked to send back the completed questionnaires in a stamped envelope preaddressed to the researcher.

Data for this study were collected using questionnaires distributed to 180 full-time employees who work in many functional areas located in South Korea. Data were collected from employees of 13 companies to examine work-family conflict across different jobs. Various industry sectors were represented, including manufacturing, finance, service, and consulting. Except for the organizations in consulting and service sectors, all of the organizations were large, publicly-traded companies. The varieties of size and industry type will increase the generalizability and external validity (Cook & Campbell, 1979). Of the 180 questionnaires, 169 were returned, for a response rate of 93.89%. Responses of 159 employees remained after removing missing data.

Of the 159 participants, 45% were male and 35% were married. Their average age was 32.8 yr. ($SD = 8.91$). Many respondents (34%) were responsible for at least one child at home. The education level of participants was diverse: 8.2% of the sample consisted of high school graduates, 19.5% had a 2-yr. college education, 57.9% had finished undergraduate university, and 14.5% had completed graduate school. Concerning their assigned jobs in the organization, 45.3% were in general management, 7.5% were in sales, 4.4% were in production, 4.4% were in research and development, 29.6% were professionals, and 8.8% worked in other areas.

Measures

In the survey, all scale items were translated into Korean and back-translated by two bilingual (English-Korean) speakers to ensure semantic equivalence (Brislin, 1980). The focal employees were asked to provide information on the proposed model. All of the items were rated on a seven-point Likert-type scale (ranging from 1: Strongly disagree to 7: Strongly agree).

Supervisor support.—Seven items from Tsui, Pearce, Porter, and Tripoli (1997)'s Supervisory Support Scale and one item, which asked directly for the employee's perceptions of his or her supervisory support, were used to assess supervisor support. A sample item reads, "My supervisor is considerate of subordinates' feelings," and the added item was: "My supervisor is supportive of me." The internal consistency reliability (Cronbach's α) of this scale was .91.

Family support.—A six-item family support scale from King, Mattimore, King, and Adams (1995) was used to obtain self-assessment of family support from the employee. A sample item is, "Someone in my family helps me out by running errands when necessary." Cronbach's α for this scale was .90.

Work-family conflict.—The 10 items developed by Netemeyer, Boles, and McMurrian (1996) were used to measure work-family conflict. These items included five items tapping work interference with family (which is

a form of inter-role conflict in which the general demands of, time devoted to, and strain created by the job interfere with performing family-related responsibilities) and five items tapping family interference with work (which is a form of inter-role conflict in which the general demands of, time devoted to, and strain created by the family interfere with performing work-related responsibilities). Hypotheses were examined in terms of both dimensions of work-family conflict. Representative items include: "The demands of my work interfere with my home and family life," "Family-related strain interferes with my ability to perform job-related duties." The reliabilities of these two scales were .95 and .94 for the work interference with family and family interference with work scales, respectively.

Emotional exhaustion.—Emotional exhaustion was assessed with nine items of Maslach and Jackson's (1981) Emotional Exhaustion Scale. The measure assesses how often respondents feel the symptoms of emotional exhaustion at work. A sample item is "I feel emotionally drained from my work." The reliability of this scale was .93.

Demographics.—To rule out alternative explanations for the relationship between the examined variables, demographic variables (i.e., age, gender, marital status, number of children, education, job position, job type, employment type, and industry type) were used as control variables. Gender was measured as a dichotomous variable coded as 1 for male and 2 for female. Education was coded as "high school" (1); "2-year college" (2); "4-year university" (3); "master's or higher degree" (4). Job position was coded as "entry level employee" (1); "supervisory" (2); "lower management" (3); "middle management" (4); "upper management or higher position" (5).

RESULTS

The means, standard deviations, reliabilities, and intercorrelations among the key variables are presented in Table 1. As suggested in this table, the correlations for most of the variables were in the expected direction.

To gauge the model fit, this study conducted a confirmatory factor analysis (CFA) and compared the model that had five latent factors (i.e., supervisor support, family support, work interference with family, family interference with work, and emotional exhaustion) with other alternative models. As shown in Table 2, the results of model comparisons indicated that the hypothesized model comprising five factors had the best fit (Arbuckle, 1997). In order to test whether the work interference with family and family interference with work are distinct constructs, they were combined into a four-factor model. Followed by this four-factor model, supervisor support and family support were combined in a three-factor model. Next, all independent variables were included in a two-factor model. Lastly, all variables were combined in a one-factor model. As summarized in Table 2, the chi-square difference test and multiple indexes

TABLE 1
DESCRIPTIVE STATISTICS OF THE VARIABLES^a

Variable ^b	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age	32.81	8.91														
2. Gender	1.55	.50	-.21†													
3. Marital status	1.65	.48	-.54‡	.26†												
4. Number of children	.52	.83	.69‡	-.23†	-.64‡											
5. Education	2.79	.79	-.02	-.10	-.03	-.01										
6. Job position	1.61	1.07	.59‡	-.35‡	-.38‡	.50‡	.25†									
7. Job type	2.92	1.98	.17	.10	.02	.01	-.03	-.02								
8. Employment type	1.36	.66	-.01	.13	.11	.07	-.07	-.08	.08							
9. Industry type	3.20	1.14	.08	.08	.00	-.04	.07	-.05	.41‡	.08						
10. Supervisor support	4.85	.89	-.06	.00	.04	.05	.11	-.04	.01	-.06	-.05	(.91)				
11. Family support	5.52	.99	-.24†	.12	.21†	-.22†	.21†	-.10	.06	-.01	-.02	.33‡	(.90)			
12. WIF	3.34	1.35	.14	-.14	-.15	.11	.22†	.20*	.02	-.18*	.03	-.22†	-.22†	(.95)		
13. FIW	2.62	1.13	.14	-.16*	-.14	.13	.11	.15	-.01	-.17*	-.09	-.33‡	-.33‡	.68‡	(.94)	
14. Emotional exhaustion	3.66	1.26	-.06	.18*	.12	-.12	.00	-.05	.00	-.04	.06	-.45‡	-.18*	.45‡	.32‡	(.93)

Note.—*N* = 159. Reliabilities are on the diagonal in parentheses. ^aMeans and standard deviations are listed for informational purposes only because these were standardized for the regression analyses. ^bThese variables were measured from focal employees. WIF = Work Interference with Family; FIW = Family Interference with Work. **p* < .05. †*p* < .01. ‡*p* < .001 (two-tailed).

(CFI, TLI, and RMSEA) all indicated that the five-factor model showed the best fit compared to other alternative models. The hypothesized model indicates that work interference with family, family interference with work, supervisor support, family support, and emotional exhaustion are separate constructs.

TABLE 2
COMPARISON OF MEASUREMENT MODELS

Model	Description	χ^2	df	CFI	TLI	RMSEA	Change from Hypothesized Model	
							$\Delta \chi^2$	Δdf
Hypothesized model	Five-factor model ^a	831.95	475	.93	.92	.07		
Model 1	Four-factor model ^b	1,083.93	479	.87	.86	.09	251.98‡	4
Model 2	Three-factor model ^c	1,596.41	482	.77	.75	.12	764.46‡	7
Model 3	Two-factor model ^d	2,242.12	484	.64	.60	.15	1,410.17‡	9
Model 4	One-factor model ^e	2,922.54	485	.50	.45	.18	2,090.59‡	10

Note.—CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = root mean square error of approximation. ^aFive-factor model: Supervisor support; Family support; Work Interference with Family; Family Interference with Work; Emotional Exhaustion. ^bFour-factor model: Supervisor support; Family support; Work Interference with Family and Family Interference with Work combined; Emotional Exhaustion. ^cThree-factor model: Supervisor support and Family support combined; Work Interference with Family and Family Interference with Work combined; Emotional Exhaustion. ^dTwo-factor model: Supervisor support, Family support combined, Work Interference with Family, and Family Interference with Work combined; Emotional Exhaustion. ^eOne-factor model: Supervisor support, Family support combined, Work Interference with Family, Family Interference with Work combined, and Emotional Exhaustion combined. ‡ $p < .001$.

Hypothesis Testing

Hierarchical regression analysis was used to test all of the hypothesized relationships. In addition, to test the indirect effect *ab* for significance, the Sobel test and bootstrapping method were conducted. Tables 3 and 4 provide a summary of the models and results used to test hypotheses for social support, work-family conflict, and emotional exhaustion. To test the hypotheses regarding the mediating role of work-family conflict, the approach suggested by Baron and Kenny (1986) was adopted. This mediation test has several important features. Firstly, the independent variable should be significantly related to the dependent variable. Secondly, the independent variable should have a significant relationship with the mediator, and finally, the mediator should be significantly related

TABLE 3
RESULTS OF MEDIATION ANALYSIS

Direct and Total Effect ^a	WIF		Emotional Exhaustion			
Variable ^b	Model 1	Model 2	Model 3	Model 4	Model 5	
Step 1: Control variables						
Age	.05	.01	.04	−.03	−.03	
Gender	−.06	−.06	.17*	.17*	.19†	
Marital status	−.06	−.03	.07	.12	.13	
Number of children	.01	.08	−.08	.04	.01	
Education	.19*	.23†	.00	.07	−.03	
Job position	.06	.04	.05	.01	−.01	
Job type	.04	.05	−.04	−.01	−.03	
Employment type	−.15	−.17*	−.06	−.10	−.03	
Industry type	.01	.00	.06	.03	.03	
Step 2: Main effect						
Supervisor support		−.26†		−.47‡	−.36‡	
Step 3: Main effect						
WIF					.42‡	
Overall F	2.00*	3.04†	.85	5.03‡	8.95‡	
R ²	.11	.17	.05	.25	.40	
F change		11.25†		40.62‡	36.19‡	
R ² change		.06		.20	.15	
Value	SE	LL 95%CI	UL 95%CI	z	p	
Indirect effect and significance using normal distribution						
Sobel	−.11	.05	−.21	−.02	−2.48	.013
Bootstrap results for indirect effect						
Effect	−.11	.05	−.22	−.03		

Note.—N = 159. ^aEntries are standardized regression coefficients. ^bVariables are standardized variables. WIF = Work Interference with Family; LL = lower limit; CI = confidence interval; UL = upper limit. Bootstrap sample size=10,000. *p < .05; †p < .01; ‡p < .001 (two-tailed).

to the dependent variables with the independent variables included in the equation. If the first three conditions hold, at least partial mediation is present. If the independent variables have non-significant beta weights in the third step, then complete mediation is present.

Hypothesis 1 posited that work interference with family mediates the relationship between supervisor support and emotional exhaustion. After all the control variables were entered, the results indicate that supervisor support is related significantly to emotional exhaustion (Model 4 of Table 3); thus, the first requirement for mediation is fulfilled. The results indicate that supervisor support is related positively to work interfer-

ence with family, as shown in Model 2 of Table 3, which fulfils the second requirement for mediation. To test the third criterion for mediation, the dependent variable (employees' emotional exhaustion) was regressed on the mediating variable (work interference with family), controlling for supervisor support. As reported in Model 5 of Table 3, the beta coefficient for work interference with family was statistically significant, reducing the coefficient of the effect of supervisor support on employees' emotional exhaustion. Therefore, the mediation analysis suggests that the effect of supervisor support on employees' emotional exhaustion is partially mediated by work interference with family.

In addition, the significance of the indirect effects were assessed by using the Sobel test and bootstrapping. The lower part of Table 3 shows that supervisor support had a negative and indirect effect on emotional exhaustion. The results of the Sobel test showed that this indirect effect was significant. Table 3 also shows that the results of the bootstrap confirmed the Sobel test. Hypothesis 1 was supported.

Family support was significantly related to emotional exhaustion (Model 4 of Table 4) and negatively related to family interference with work (Model 2 of Table 4), thus the first and second requirements for mediation were fulfilled. To test the third criterion for mediation, the dependent variable (employees' emotional exhaustion) was regressed on the mediating variable (family interference with work), controlling for family support. As reported in Model 5 of Table 4, the beta coefficient for family interference with work was significant, reducing the coefficient of the effect of family support on employees' emotional exhaustion. Therefore, the effect of family support on employees' emotional exhaustion was fully mediated by family interference with work. Also, the lower part of Table 4 shows that family support had a negative and indirect effect on emotional exhaustion. The results of the Sobel test showed that this indirect effect was significant. Hypothesis 2 was supported.

DISCUSSION

The main purpose of this research was to examine whether work-family conflict mediates the relationship between social support and emotional exhaustion based on COR theory. The results suggested that the relationship between supervisor support and emotional exhaustion was mediated by work interference with family, while the relationship between family support and emotional exhaustion was mediated by family interference with work.

Theoretical Implications

The results provide an integrative framework on social support, work-family conflict, and emotional exhaustion by investigating the

TABLE 4
RESULTS OF MEDIATION ANALYSIS

Direct and Total Effect ^a		FIW		Emotional Exhaustion		
Variable ^b		Model 1	Model 2	Model 3	Model 4	Model 5
Step 1: Control variables						
Age		.08	.02	.04	-.01	-.01
Gender		-.09	-.07	.17*	.19*	.22†
Marital status		-.03	.00	.07	.09	.09
Number of children		.04	.02	-.08	-.09	-.10
Education		.10	.17*	.00	.05	.00
Job position		.00	.01	.05	.06	.06
Job type		.04	.09	-.04	-.01	-.04
Employment type		-.15	-.15*	-.06	-.06	-.01
Industry type		-.10	-.13	.06	.04	.09
Step 2: Main effect						
Family support			-.36‡		-.25†	-.13
Step 3: Main effect						
FIW						.34‡
Overall <i>F</i>		1.43	3.45‡	.85	1.72‡	3.30‡
<i>R</i> ²		.08	.19	.05	.10	.20
<i>F</i> change			20.03‡		9.13†	17.27‡
<i>R</i> ² change			.11		.06	.09
	Value	SE	LL 95% CI	UL 95% CI	<i>z</i>	<i>p</i>
Indirect effect and significance using normal distribution						
Sobel	-.12	.04	-.21	-.04	-2.78	.005
Bootstrap results for indirect effect						
Effect	-.12	.05	-.25	-.05		

Note.—*N* = 159. ^aEntries are standardized regression coefficients. ^bVariables are standardized variables. FIW = Family Interference with Work; LL = lower limit; CI = confidence interval; UL = upper limit. Bootstrap sample size = 10,000. **p* < .05; †*p* < .01; ‡*p* < .001 (two-tailed).

mediating effect of work-family conflict in the relation between social support and employees' emotional exhaustion. Specifically, supervisor and family support, respectively, had a buffering or reducing effect on work interference with family and family interference with work within the same domains (Aryee, Luk, Leung, & Lo, 1999; Nielson, Carlson, & Lankau, 2001), which was related to lower emotional exhaustion. Furthermore, each aspect of work-family conflict seems to be a separate mechanism between support originating from different domains and emotional exhaustion. To date, most prior research on work-family conflict has focused on only its antecedents or consequences, and thus there

is still no deep understanding of the process inter-relating social support, work-family conflict, and emotional strain. This study contributes to the work-family conflict literature by suggesting an integrated perspective of work-family conflict.

Work-family conflict was bidirectionally examined, i.e., in terms of work interference with family and family interference with work. A recent literature review of work-family conflict suggested a "matching hypothesis" that work interference with family should have stronger effects on work-related outcomes and family interference with work should have stronger effects on family-related outcomes (Amstad, *et al.*, 2011). In line with this notion, the relation between supervisor support and emotional exhaustion was mediated by work interference with family, and the relation between family support and emotional exhaustion was mediated by family interference with work, in the matching domains. Therefore, this study provided additional empirical evidence supporting the matching hypothesis along with the recent literature of work-family conflict.

Lastly, work-family conflict was examined using a sample of employees in South Korea. In this country, there is a large female participation in the work force, combined with conservative gender roles. Since cultures influence gender role expectations as well as the extent to which work and family are perceived as distinct domains (Aryee, *et al.*, 1999), it is essential to investigate work-family conflict in non-Western societies, especially considering the importance of Asian economies. The present study, therefore, contributes to the growing cross-cultural literature on work-family conflict in the context of Asian cultures.

Practical Implications

From a practical perspective, the noteworthy finding of this study is the importance of supervisor support as an important factor restricting employees' work-family conflict and their emotional exhaustion in the work setting. As mentioned above, supervisors are the key people who have a powerful effect on subordinates' organizational attitude and behaviors (Beehr, Farmer, Glazer, Gudanowski, & Nair, 2003). Moreover, according to Beehr (1995), social support from work-related sources is more important in the job stress process than non-work related support. Although many organizations use family-responsive policies, such as day-care centers, maternity and paternity leave, and alternative work schedules (Zedeck & Mosier, 1990; Russell, O'Connell, & McGinnity, 2009), a supportive supervisor can more directly help employees to manage their work and family issues. Thus, this study suggests that organizations should simultaneously attend not only to implementation of family-supportive policies but also to train supervisors to be supportive of their employees.

In addition, examining the direct effect of employees' work-family conflict on emotional exhaustion, the findings suggest a general implication for organizations. A number of earlier studies have suggested that emotional exhaustion has a negative influence not only on employee job performance but also on organizational effectiveness (Cordes & Dougherty, 1993; Demerouti, *et al.*, 2001; Witt, Andrews, & Carlson, 2004). For example, at an individual level, emotional exhaustion is related to the experience of anxiety, distress, and reduced self-esteem (Maslach, Schaufeli, & Leiter, 2001). Furthermore, previous studies have suggested that emotional exhaustion is negatively related to job satisfaction and task performance, but positively associated with turnover intentions (Firth & Britton, 1989; Wright & Bonett, 1997). Thus, managing employees' emotional exhaustion may be a necessary and useful practice for organizational effectiveness. In line with this issue, organizations should not overlook employees' conflict between work and family as sources of their burnout, especially emotional exhaustion. Managing work-family conflicts may reduce the emotional exhaustion associated with undesirable outcomes for both individuals and organizations.

Limitations and Directions for Future Research

There are limitations to consider in the interpretation of the results. Firstly, by collecting only self-report data, there is a risk of common method bias. However, the purpose of this study permits self-reports (Spector, 1994), and the presence of common method bias does not necessarily affect results or conclusions (Spector, 2006). Nevertheless, it is valuable to use multiple informants to minimize common method bias in the future. Secondly, given the cross-sectional design of this study, causal conclusions cannot be drawn. For example, it is equally reasonable to suppose that emotional exhaustion might increase work-family conflict and lead employees to perceive low social support. The causal directions of relations among social support, work-family conflict, and emotional exhaustion were interpreted in line with prior research (Boles, *et al.*, 1997; Taylor, Delcampo, & Blancero, 2009) and theory, but doing so does not rule out the possibility of the opposite causality. Therefore, it is worthwhile to use longitudinal design to provide a firm evidence of causation in the future.

REFERENCES

- ADAMS, G. A., KING, L. A., & KING, D. W. (1996) Relationships of job and family involvement, family social support, and work-family conflict with job and life satisfaction. *Journal of Applied Psychology*, 81, 411-420.
- AMSTAD, F. T., MEIER, L. L., FASEL, U., ELFERING, A., & SEMMER, N. K. (2011) A meta-analysis of work-family conflict and various outcomes with a special emphasis on cross-domain versus matching-domain relations. *Journal of Occupational Health Psychology*, 16, 151-169.

- ARBUCKLE, J. L. (1997). *Amos users' guide*. Chicago, IL: Small Waters Corporation.
- ARYEE, S., LUK, V., LEUNG, A., & LO, S. (1999) Role stressors, interrole conflict, and well-being: the moderating influence of spousal support and coping behaviors among employed parents in Hong Kong. *Journal of Vocational Behavior*, 54, 259-278.
- BACHARACH, S. B., BAMBERGER, P., & BIRON, M. (2010) Alcohol consumption and workplace absenteeism: the moderating effect of social support. *Journal of Applied Psychology*, 95, 334-348.
- BALLOUT, H. I. (2008) Work-family conflict and career success: the effects of domain-specific determinants. *Journal of Management Development*, 27, 437-466.
- BARON, R. M., & KENNY, D. A. (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- BEEHR, T. A. (1995) *Psychological stress in the workplace*. New York: Routledge.
- BEEHR, T. A., FARMER, S. J., GLAZER, S., GUDANOWSKI, D. M., & NAIR, V. N. (2003) The enigma of social support and occupational stress: source congruence and gender role effects. *Journal of Occupational Health Psychology*, 8, 220-231.
- BERNAS, K. H., & MAJOR, D. A. (2000) Contributors to stress resistance: testing a model of women's work-family conflict. *Psychology of Women Quarterly*, 24, 170-178.
- BOLES, J. S., JOHNSTON, M. W., & HAIR, J. F., JR. (1997) Role stress, work-family conflict and emotional exhaustion: inter-relationships and effects on some work-related consequences. *Journal of Personal Selling and Sales Management*, 17, 17-28.
- BRISLIN, R. W. (1980) Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology*. Boston, MA: Allyn & Bacon. Pp. 349-444.
- BYRON, K. A. (2005) A meta-analytic review of work-family conflict and its antecedents. *Journal of Vocational Behavior*, 67, 169-198.
- CARLSON, D. S., & PERREWÉ, P. L. (1999) The role of social support in the stressor-strain relationship: an examination of work-family conflict. *Journal of Management*, 25, 513-540.
- COOK, T. D., & CAMPBELL, D. T. (1979) *Quasi-experimentation: design and analysis for field settings*. Boston, MA: Houghton Mifflin.
- CORDES, C. L., & DOUGHERTY, T. W. (1993) A review and an integration of research on job burnout. *Academy of Management Review*, 18, 621-656.
- CROPANZANO, R., RUPP, D. E., & BYRNE, Z. S. (2003) The relationship of emotional exhaustion to work attitudes, job performance, and organizational citizenship behaviors. *Journal of Applied Psychology*, 88, 160-169.
- DEMEROUTI, E., BAKKER, A. B., NACHREINER, F., & SCHAUFELI, W. B. (2001) The job demands-resources model of burnout. *Journal of Applied Psychology*, 86, 499-512.
- DEMEROUTI, E., GEURTS, S. A. E., & KOMPIER, M. (2004) Positive and negative work-home interaction: prevalence and correlates. *Equal Opportunities International*, 23, 6-35.
- FIRTH, H., & BRITTON, P. (1989) 'Burnout,' absence and turnover amongst British nursing staff. *Journal of Occupational Psychology*, 62, 55-59.
- FRONE, M. R. (2003) Work-family balance. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology*. Washington, DC: American Psychological Association. Pp. 143-162.
- GRANDEY, A. A., & CROPANZANO, R. (1999) The conservation of resources model applied to work-family conflict and strain. *Journal of Vocational Behavior*, 54, 350-370.

- GREENHAUS, J. H., & BEUTELL, N. J. (1985) Source of conflict between work and family roles. *Academy of Management Review*, 10, 76-88.
- GUTTEK, B. A., SEARLE, S., & KLEPA, L. (1991) Rational versus gender role explanations for work-family conflict. *Journal of Applied Psychology*, 76, 560-568.
- HALBESLEBEN, J. R. B. (2006) Sources of social support and burnout: a meta-analytic test of the conservation of resources model. *Journal of Applied Psychology*, 91, 1134-1145.
- HOBFOLL, S. (1989) Conservation of resources: a new attempt at conceptualizing stress. *American Psychologist*, 44, 513-524.
- KING, L. A., MATTIMORE, L. K., KING, D. W., & ADAMS, G. A. (1995) Family support inventory for workers: a new measure of perceived social support from family members. *Journal of Organizational Behavior*, 16, 235-258.
- LAPIERRE, L. M., & ALLEN, T. D. (2006) Work-supportive family, family-supportive supervision, use of organizational benefits, and problem-focused coping: implications for work-family conflict and employee well-being. *Journal of Occupational Health Psychology*, 11, 169-181.
- LIVINGSTON, B. A., & JUDGE, T. A. (2008) Emotional responses to work-family conflict: an examination of gender role orientation among working men and women. *Journal of Applied Psychology*, 93, 207-216.
- MASLACH, C. (1982) *Burnout: the cost of caring*. Englewood Cliffs, NJ: Prentice Hall.
- MASLACH, C., & JACKSON, S. E. (1981) The measurement of experienced burnout. *Journal of Occupational Behavior*, 2, 99-113.
- MASLACH, C., SCHAUFELI, W. B., & LEITER, M. P. (2001) Job burnout. *Annual Review of Psychology*, 52, 397-422.
- MESMER-MAGNUS, J. R., & VISWESVARAN, C. (2005) Convergence between measures of work-to-family and family-to-work conflict: a meta-analytic examination. *Journal of Vocational Behavior*, 67, 215-232.
- MICHEL, J. S., KOTRBA, L. M., MITCHELSON, J. K., CLARK, M. A., & BALTES, B. B. (2011) Antecedents of work-family conflict: a meta-analytic review. *Journal of Organizational Behavior*, 32, 689-725.
- MICHEL, J. S., MITCHELSON, J. K., PICHLER, S., & CULLEN, K. L. (2010) Clarifying relationships among work and family social support, stressors, and work-family conflict. *Journal of Vocational Behavior*, 76, 91-106.
- NASURDIN, A. M., & HSIA, K. L. (2008) The influence of support at work and home on work-family conflict: does gender make a difference? *Research and Practice in Human Resource Management*, 16, 18-38.
- NETEMEYER, R. G., BOLES, J. S., & MCMURRIAN, R. (1996) Development and validation of work-family conflict and family-work conflict scales. *Journal of Applied Psychology*, 81, 400-410.
- NIELSON, T. R., CARLSON, D. S., & LANKAU, M. J. (2001) The supportive mentor as a means of reducing work-family conflict. *Journal of Vocational Behavior*, 59, 364-381.
- ODLE-DUSSEAU, H. N., BRITT, T. W., & GREENE-SHORTTRIDGE, T. M. (2012) Organizational work-family resources as predictors of job performance and attitudes: the process of work-family conflict and enrichment. *Journal of Occupational Health Psychology*, 17, 28-40.
- RUSSELL, H., O'CONNELL, P. J., & MCGINNITY, F. (2009) The impact of flexible working arrangements on work-life conflict and work pressure in Ireland. *Gender, Work, & Organization*, 16, 73-97.

- SHIROM, A. (2003) Job-related burnout: a review. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology*. Washington, DC: American Psychological Association. Pp. 245-264.
- SPECTOR, P. E. (1994) Using self-report questionnaires in OB research: a comment on the use of a controversial method. *Journal of Organizational Behavior*, 15, 385-392.
- SPECTOR, P. E. (2006) Method variance in organizational research: truth or urban legend? *Organizational Research Methods*, 9, 221-232.
- TAYLOR, B. L., DELCAMPO, R. G., & BLANCERO, D. M. (2009) Work-family conflict/facilitation and the role of workplace supports for U.S. Hispanic professionals. *Journal of Organizational Behavior*, 30, 643-664.
- TSUI, A. S., PEARCE, J. L., PORTER, L. W., & TRIPOLI, A. M. (1997) Alternative approaches to the employee-organization relationship: does investment in employees pay off? *Academy of Management Journal*, 40, 1089-1121.
- VAN YPEREN, N. W., & HAGEDOORN, M. (2003) Do high job demands increase intrinsic motivation or fatigue or both? The role of job control and job social support. *Academy of Management Journal*, 46, 339-348.
- WITT, L. A., ANDREWS, M. C., & CARLSON, D. S. (2004) When conscientiousness isn't enough: emotional exhaustion and performance among call center customer service representatives. *Journal of Management*, 30, 149-160.
- WRIGHT, T. A., & BONETT, D. G. (1997) The contribution of burnout to work performance. *Journal of Organizational Behavior*, 18, 491-499.
- YU, M-C., LEE, Y-D., & TSAL, B-C. (2010) Relationships among stressors, work-family conflict, and emotional exhaustion: a study of electronics industry employees in China. *Social Behavior and Personality*, 38, 829-844.
- ZEDECK, S., & MOSIER, K. L. (1990) Work in the family and employing organization. *American Psychologist*, 45, 240-251.

Accepted September 20, 2013.

APPENDIX 1

Supervisor support

1. My supervisor seems willing to listen to my problems.
2. My supervisor is considerate of subordinates' feelings.
3. I can rely on my supervisor.
4. My supervisor is friendly.
5. I don't really feel that my supervisor and I are working toward shared or team objectives (R).
6. My supervisor seems to be rather distant and unapproachable (R).
7. Supervisors and subordinates have confidence in one another.
8. My supervisor is supportive of me.

Family support

1. Members of my family want me to enjoy my job.
2. Members of my family are happy for me when I am successful at work.
3. When I succeed at work, members of my family show that they are proud of me.

4. Someone in my family helps me out by running errands when necessary.
5. Members of my family are willing to straighten up the house when it needs it.
6. Members of my family cooperate with me to get things done around the house.

Work interference with family

1. The demands of my work interfere with my home and family life.
2. The amount of time my job takes up makes it difficult to fulfill family responsibilities.
3. Things I want to do at home do not get done because of the demands my job puts on me.
4. My job produces strain that makes it difficult to fulfill family duties.
5. Due to work-related duties, I have to make changes to my plans for family activities.

Family interference with work

1. The demands of my family or spouse/partner interfere with work-related activities.
2. I have to put off doing things at work because of demands on my time at home.
3. Things I want to do at work don't get done because of the demands of my family or spouse/partner.
4. My home life interferes with my responsibilities at work such as getting to work on time, accomplishing daily tasks, and working overtime.
5. Family-related strain interferes with my ability to perform job-related duties.

Emotional exhaustion

1. I feel emotionally drained from my work.
2. I feel used up at the end of the workday.
3. I feel fatigued when I get up in the morning and have to face another day on the job.
4. Working with people all day is really a strain for me.
5. I feel burned out from my work.
6. I feel frustrated by my job.
7. I feel I'm working too hard on my job.
8. Working with people directly puts too much stress on me.
9. I feel like I'm at the end of my rope.

PSYCHOLOGICAL REPORTS

ISSN 0033-2941 OCTOBER 2013

*Sociocultural
Issues in
Psychology*



PEOPLE'S PREFERENCE PATTERNS FOR GAINS/LOSSES IN MULTIPLE TIME PERIOD SITUATIONS¹

SHIN-SHIN CHANG

*Department of Business Administration
Chung Yuan Christian University, Taiwan*

JUNG-HUA CHANG

*Department of Marketing and Logistics
MingDao University, Taiwan*

Summary.—Little research to date has been devoted to investigating whether people treat time differently from money when facing multiple gains or losses. This study tested the hypothesis that because time is characterized by perishability, fixed supply, and infungibility, people with strong motivation to obtain a long period of uninterrupted discretionary time would strive to trim the time needed for non-discretionary activities or to combine several non-discretionary activities. As a result, people prefer integration over segregation of multiple time losses or gains, which is not consistent with the prediction based on hedonic editing theory or the renewable resource model. This proposition is supported by results from four experiments.

If a girl receives \$20 from her father in the morning and then another \$20 from her mother in the evening, she is predicted to be happier than she would have been had she received \$40 at the same time (Thaler, 1985). According to hedonic editing theory, people prefer to segregate multiple monetary gains and integrate multiple monetary losses to maximize their pleasure (Thaler, 1985; Thaler, 1999). That is, the segregation of monetary gains is more valuable than integration. However, when people face decisions involving multiple gains and losses of time, will they behave in the same way as they do for monetary gains or losses? For example, will an employee prefer to take a 15-min. break in the morning and another 15-min. break in the afternoon over taking a 30-min. break in a working day? The question is whether people prefer to segregate several time gains just as they prefer to segregate monetary gains. What are the underlying mechanisms for people's decisions about gains and losses of multiple periods of time?

A growing body of research has suggested that people treat time differently than they do money. Their decisions regarding how to spend their time are more heuristic than those about spending money (Saini & Monga, 2008). When tracking the costs of paying, for example, people usually write

¹Address correspondence to Shin-Shin Chang, 200 Chung Pei Rd., Chung Li, Taiwan or e-mail (sschang351@cycu.edu.tw).

off the time costs at the end of the accounting period, but they write off the monetary costs across accounting periods (Soster, Monga, & Bearden, 2010). Moreover, people are more likely to underestimate the value of time (Leclerc, Schmitt, & Dubé, 1995; Okada & Hoch, 2004; Chang, Chang, Chang, & Chien, 2013) and ignore the sunk costs (Soman, 2001). Therefore, it seems plausible to conclude that people's decisions about multiple time gains or losses may be different from those about monetary outcomes.

Building on mental accounting and prospect theory, Leclerc, *et al.* (1995) investigated whether people treat time like money when they make waiting-time decisions, which, by their definition, are time-loss situations. They suggest that, under deterministic conditions, people make time-loss decisions just like those for money, in accordance with the convex loss function proposed by prospect theory. However, in risky situations, people demonstrate risk-averse behavior, which is not consistent with the predicted risk-taking behavior regarding decisions involving losses of money. In particular, people prefer to have time losses integrated when these losses are already combined to them as a time lump-sum cost. Besides, given a choice about the timing of events, people are also inclined to integrate waiting time. In sum, Leclerc, *et al.* (1995) suggested there is a strong tendency toward the integration of time losses, which opposes the assertion that people prefer separate losses (Thaler & Johnson, 1990) and are prone to experience small losses on separate days (Linville & Fischer, 1991). Researchers have presented three explanations for this discrepancy. Firstly, waiting may not be coded as a loss, but as a cost: waiting may only be coded as a loss when actual waiting time exceeds expected waiting time. Secondly, the loss function of time may not be convex, but instead, concave. Thirdly, because time is less transferable than money, people may prefer an uninterrupted half-day or a half-day interrupted by a longer period to two half-days interrupted by errands.

In this study, an alternative explanation of the distinction between people's decisions regarding time-loss and money-loss events is proposed. People may prefer integration over segregation of multiple time losses mainly because they tend to want more discretionary time. The term "discretionary time" is defined in this study as "the amount of time people can spend or use at their own will." A person's discretionary time in a day is 24 hours minus the time needed for necessary activities, such as work, sleeping, etc. It is presumed that most people like to have the freedom to arrange their time as they choose. This research posits that the perishability, infungibility, and fixed-supply characteristics of time, as well as people's inclination to maximize their amounts of discretionary time, will lead to the preference of integration over segregation of multiple time losses and time gains.

Theoretical Background

Time is a unique resource. Time is a fixed-supply source because there are only 24 hours in a day and seven days in a week (Feldman & Hornik, 1981; Gross, 1987; Okada & Hoch, 2004). Unlike money, time cannot be earned. People can only allocate or reallocate their time between daily activities. For example, gains in sleeping time may come from losses of discretionary time. Time is also perishable and infungible. Time cannot be stored for future use or borrowed from the future for immediate spending (Gross, 1987; Leclerc, *et al.*, 1995; Soman, 2001; Zauberger & Lynch, 2005; Soster, *et al.*, 2010). Most people perceive time to be linear and separable, a one-direction road that stretches from the past into the future (Feldman & Hornik, 1981; Graham, 1981; Soster, *et al.*, 2010).

Monetary or financial gains and losses are usually defined with respect to some reference point, such as people's current assets (Kahneman & Tversky, 1979). They are related to the changes in the status quo rather than absolute wealth (Thaler, 1985; Thaler, 1999). By the same token, it is argued that temporal gains and losses can only be defined with reference points. For example, the time loss in waiting for buying a theater ticket is the amount of time that exceeds the expected time (Leclerc, *et al.*, 1995).

However, time losses associated with waiting may be interpreted differently when the concepts of discretionary and non-discretionary time are introduced to this context. Whereas non-discretionary time is time spent on necessary activities, discretionary time is a person's available time minus their non-discretionary time. In general, time for paid tasks is coded as non-discretionary, while time for vacation or leisure activities is regarded as discretionary. If waiting is perceived as non-discretionary time, the waiting itself is a time loss with regard to a person's discretionary time. That is, the total time in waiting, not merely the time that exceeds the expected waiting time, is perceived as a loss of discretionary time. Hence, the gains and losses of time in this study are defined as an increase or decrease in a person's discretionary time.

Multiple Gains and Losses Involving Money and Time

Based on prospect theory (Kahneman & Tversky, 1979), the hedonic editing theory argues that people frame multiple outcomes to maximize their pleasure by (a) segregating gains; (b) integrating losses; (c) integrating small losses with larger gains; and (d) segregating smaller gains with larger losses (called the "silver lining") to preserve their hedonic feelings (Thaler, 1985). When people have two or more monetary gains simultaneously, given that the value function of monetary gains is concave (hedonic editing), those two separate gains are likely to bring more value (pleasure). On the contrary, owing to the convexity of the loss function,

the combination of two monetary losses results in a smaller decrease of utility than would segregation. However, some research has suggested that people prefer an integration of losses only when the prospects are already combined for them. They may favor a segregation of losses when they are given a choice in regards to the timing of events (Thaler & Johnson, 1990).

The renewable resources model provides an alternative explanation of monetary outcomes by assuming that people spend limited but renewable resources, such as cognitive resources, social resources, and physical energy, to deal with the decisions of multiple events. Specifically, people use "limited gain-savoring resources" to process information about multiple gains and "loss-buffering resources" to process information about multiple losses (Linville & Fischer, 1991). Accordingly, whether people prefer to integrate or separate their monetary gains and losses depends on their psychological resources. Related empirical studies have shown that people prefer to segregate two financial losses only for large losses. People have equal preferences for the integration or segregation of small losses (Linville & Fischer, 1991). In sum, people prefer integrated monetary losses when the losses are already combined for them, but they favor segregation of monetary losses when they are given a choice in regard to the timing of events.

If people treat temporal and financial decisions similarly, it is predicted that they should favor an integration of temporal losses when these losses are combined for them. They are likely to prefer a segregation of temporal losses otherwise. However, Leclerc, *et al.* (1995) have shown that people consistently display strong tendencies to prefer integration of temporal losses in waiting situations. They attribute this conflicting finding to the infungibility of time, the concavity of the loss function for time, and that waiting is not coded as a loss, but as a cost.

Unlike Leclerc, *et al.* (1995), who studied on decisions about multiple time losses, this research is aimed at exploring decisions with respect to both time losses and gains. Based on the concept of discretionary time, this research posits that people are strongly motivated to maximize their available discretionary time in a day or in a certain period of time because they usually enjoy discretionary activity more. They can do this by reducing the time spent on each non-discretionary activity. The time needed to complete non-discretionary activities is regarded as a loss of discretionary time. Suppose Mr. X waits in line to cash a check in a bank and waits a total of 30 minutes before completing the transaction. If Mr. X takes this waiting as a non-discretionary activity, he would then code this 30 minutes as a loss of time because it decreases his total discretionary time in a day. If Mr. X takes 50 minutes to finish cashing the check due to the inci-

dental breakdown of equipment, which exceeds his anticipated waiting time of 30 minutes, then he incurs a total loss of 50 minutes with respect to discretionary time. The additional 20 minutes would be regarded as an additional loss of discretionary time.

When people are not able to decrease their non-discretionary time, they may resort to re-arranging some of their non-discretionary activities. This can include completing mandatory tasks uninterrupted so as to gain a longer period of uninterrupted discretionary time later. Given a constant amount of discretionary time, people are likely to avoid interrupted non-discretionary time because it results in interrupted discretionary time. More uninterrupted discretionary time will bring more flexibility in choosing discretionary activities. Imagine a multiple time loss involving the completion of two errands in two days, and the completion of each errand will take one hour. Most people are inclined to complete these two errands consecutively to obtain a longer uninterrupted discretionary time.

To most people, a time gain is an increase in discretionary time or a longer uninterrupted discretionary time. As to the case involving multiple time gains, suppose company A announces that its sale representatives with outstanding performances in the previous quarter will be granted two days off. It is seemingly plausible that most people are likely to choose two adjacent, rather than non-adjacent, days.

There are several ideas tested in this set of studies. Firstly, people must allocate all their time to either non-discretionary or discretionary activities. Because time is a fixed-supply resource, a time loss or an increase of non-discretionary time implies a loss or decrease of discretionary time, whereas a time gain or a decrease of non-discretionary time implies a gain or increase of discretionary time. Secondly, people generally prefer discretionary time over non-discretionary time because they can allocate it as they please. As a consequence, they are more likely to interpret time losses or gains in reference to discretionary, rather than non-discretionary, time. Thirdly, to enjoy discretionary time more, people strive to obtain more discretionary time or to decrease time spent on non-discretionary activities. Fourthly, aside from people's attempts to increase discretionary time, they also tend to avoid several interrupted non-discretionary activities, because the discretionary time between two non-adjacent, non-discretionary events cannot be stored or transferred to combine with the later discretionary time. As a result, people have strong preferences for integration over segregation of multiple gains involving discretionary time or multiple losses involving non-discretionary time. This observation opposes the assertion that people tend to favor segregation over integration of multiple monetary gains.

There were four experiments. Experiment 1 explored the issue of multiple discretionary time gains with regard to whether people prefer to combine an extra discretionary time with the discretionary time at hand into one longer period. Experiments 2, 3, and 4 investigated decision contexts involving multiple non-discretionary time losses, with the goals of assessing whether people prefer integration over segregation of two separate non-discretionary time losses.

Experiment 1:

Multiple Discretionary Time Gains for Choosing One Day Off

METHOD

Participants

Experiment 1 was conducted to understand how people choose a day off. Sixty-five undergraduates (38 women, 27 men) from a university in Taiwan participated in this experiment to earn extra credit. Their mean age was 18.4 yr. ($SD = 0.6$).

Procedure

Participants in Experiment 1 were presented with a scenario in which a regular employee who worked five days per week, from Monday to Friday, had to choose a day off. After reading this scenario, the participants were instructed to select one day on a five-day table (from Monday to Friday) and then write down the reasons for their choice. People who preferred integration of multiple time gains were inclined to choose a Monday or Friday to extend their holiday. In contrast, people who preferred their time gains to be segregated chose Tuesday, Wednesday, or Thursday for their days off.

RESULTS AND DISCUSSION

Participants who selected Monday or Friday belonged to the "integrated group," whereas those who selected Tuesday, Wednesday, or Thursday belonged to the "segregated group."

Two doctoral students, who were blind to the treatment conditions, were recruited to encode the participants' responses independently. Before coding, they had to learn the definitions of the primary categories. Participant responses were categorized into one of the following codes: holiday extended (e.g., three-day vacation), blue Monday/Friday fever (e.g., not willing to work on Monday), taking break during the week (e.g., vacation on one of the five working days), others (e.g., beat the crowds), and no reasons. Inter-judge reliability (kappa coefficient of agreement) for the five categories was .80. Disagreements were resolved by discussion.

Fifty-seven participants chose to take a day off on a Monday or Friday. Thirty-six (63%) said they intended to extend their weekend while nine

(16%) said they had "blue Monday" or "Friday fever," which indicated they wanted to extend their weekend vacation. Four (7%) listed other reasons, such as "Friday is a good day," or "taking Monday off would beat the crowds." Eight (14%) left no comments.

Only eight participants intended to take a day off on Tuesday, Wednesday, or Thursday. Six of the eight participants said they could take a day off on any of the working days, and the other two gave no reason. To examine whether the proportion of the two groups were equal, a chi-square test was conducted. The result showed that approximately 88% of participants chose Monday ($n = 31$) or Friday ($n = 26$), significantly more ($\chi^2 = 36.94, p < .001$, Cramér's $\Phi = .75$) than those who chose the other three days (12%: Tuesday = 1; Wednesday = 7; Thursday = 0).

Results from Experiment 1 indicated that people were inclined to integrate their discretionary time gains. Generally speaking, workers dislike Monday because it is the first day of the work week. "Friday fever" describes the fact that many workers usually postpone their assignments to the last one or two work days, causing Friday's workloads to be larger than those on other days. These explanations clarify people's preference to take a day off on Monday or Friday.

Experiment 2:

Multiple Non-discretionary Time Losses in Waiting

METHOD

Participants

Forty-six undergraduates (26 women, 20 men) in Taiwan, ages of 19 to 24 (M age = 20.7 yr., $SD = 1.1$), were invited to participate in Experiment 2. They were all Chinese students and received an extra class credit after the experiment.

Procedure

All participants were presented with a scenario about a waiting situation. They were asked to imagine they were waiting in line for six hours and were then asked which two hours of the six they would like to choose for a break (Appendix A). Their answers were encoded as integrated or segregated depending on whether two adjacent hours were chosen or not. For example, a choice of the first two hours for the break was classified as integration of multiple time savings, whereas a choice of the first and the third hours was classified as segregation. For a six-hour period, there are 15 possible combinations of two hours. Among all combinations, only five of those are adjacent: hours 1 and 2, 2 and 3, 3 and 4, 4 and 5, and 5 and 6. The other 10 combinations are segregated.

RESULTS AND DISCUSSION

Among 46 participants, 34 indicated a choice of two adjacent hours (73.9%), and 12 chose separate hours (26.1%). The number choosing integrated breaks was significantly larger ($\chi^2 = 10.55$, $p = .001$, Cramér's $\Phi = .48$). As expected, people preferred to integrate their multiple time gains.

For participants in the integration group, 22 chose to take a break during the 3rd and 4th hours, five chose to take a break in the last two hours, and five chose to take the first two hours for a break. An explanation for these results is that most participants chose the 3rd and 4th hours because a continuous four-hour wait for a break would be more tiring. Results of Experiment 2 showed that the participants preferred to integrate two separate losses involving nondiscretionary time to obtain a long uninterrupted discretionary time period.

*Experiment 3:**Multiple Time Losses in Arrangement of Working Days*

METHOD

Participants

Sixty-three Chinese undergraduates (27 men, 36 women) from a university in Taiwan were invited to participate in this experiment. Their ages ranged from 18 to 22 (M age = 18.3 yr., $SD = 0.7$). They were given an extra course credit after the experiment.

Procedure

All participants were instructed to read a fictional scenario about the arrangement of two working days in a week. They were told to imagine that they were temporary workers on an assignment. They could select two days for assignments during the week (Appendix B). Similar to Experiment 1b, if the two working days chosen were adjacent, the response was coded as "integration." If the participant selected two separate days, it was coded as "segregation."

RESULTS AND DISCUSSION

Forty-one participants (65%) preferred to arrange two assignments on two conjunctural days (integration), and 22 (34%) participants selected two separate days for their assignments (segregation). These results indicate that people prefer to integrate their time losses ($\chi^2 = 5.73$, $p < .05$, Cramér's $\Phi = .30$).

The results showed that most participants preferred to integrate their time losses. However, some participants did choose to work on two separate days. One explanation may be that they thought it would be more difficult to work two consecutive days. Another explanation is that choosing to work two consecutive days would mean a five-day vacation, which could be too long for some.

Experiment 4:
Multiple Time Losses in Completion of Two Errands
METHOD

Participants

The participants were 45 undergraduates (17 men, 28 women) from a university in Taiwan. Their average age was 20.5 yr. ($SD = 0.6$). They were recruited from a management class and were offered an extra course credit for taking part in this experiment.

Procedure

The scenario in Experiment 4 was adapted from Leclerc, *et al.*'s Study 4 (1995), which was designed to investigate how people treat two fragmented time losses and the reasons they made their choices. Participants were required to choose between doing two errands on the same day or on different days (Appendix C). They were asked to write down their reasons for their decisions.

RESULTS AND DISCUSSION

Forty-four participants selected the option indicating they would complete both errands in one day (98%, $\chi^2 = 41.09$, $p < .001$, Cramér's $\Phi = .97$). This result is similar to those of Leclerc, *et al.*'s Study 4, in which a larger proportion of participants had a clear preference for the integration of multiple time losses. Next, two trained coders (Ph.D. students) categorized the participants' responses. They were blind to the experiment purpose and treatment conditions. The inter-judge reliability was .79 (kappa coefficient of agreement). Disagreements were solved via discussion. The reasons reported by the participants included the following: (1) one complete day, e.g., could have another complete day off; (2) convenience, e.g., two tasks completed at one time; (3) no reasons. Most participants desired to have a whole day off (42%) or thought completing two tasks at one time would be more efficient (42%). This showed that participants wished to integrate multiple time losses to have a non-fragmented discretionary time period later.

General Discussion

Based on hedonic editing theory and the renewable resources model, it is suggested that people prefer segregation of monetary gains, but prefer integration of monetary losses in general. This research posits that people may behave differently when they face decisions involving time gains or losses. Unlike Leclerc, *et al.* (1995), which focused on decisions about multiple time losses, the purpose of this research was exploring decisions with respect to both time losses and time gains. Specifically, results showed that people favor integration of multiple temporal gains or losses over

segregation. One possible explanation for the discrepancy between this research and previous research is rooted in how people associate time losses or gains with discretionary and non-discretionary time. It is argued that people tend to perceive time spent on non-discretionary activities as a loss, but tend to perceive time spent on discretionary activities as a gain. Since discretionary time is time people can arrange at their own will, they are motivated to seek more of it in their daily lives. Due to the fixed-supply nature of time, an increase in non-discretionary time surely leads to a decrease or loss in discretionary time, and a decrease in non-discretionary time results in an increase or gain in discretionary time. If people are not able to trim non-discretionary activities, they seek to combine several interrupted non-discretionary activities to get more uninterrupted discretionary time because of the perishability and infungibility of time. It is interesting to note that people may not have the same tendency or motivation to integrate two separate sums of money because two sums gained at different times can be easily added together. Consequently, there is a demonstrated preference toward the integration of time losses involving non-discretionary events and the integration of time gains involving discretionary time.

This research is contributive in its finding that people may behave differently for financial and temporal decisions. It also contributes to the theories in consumer psychology by incorporating the specific characteristics of time, such as infungibility, perishability, and its fixed-supply nature, to provide the underlying mechanisms for people's different decisions involving multiple time gains or losses.

REFERENCES

- CHANG, C-C., CHANG, S-S., CHANG, J-H., & CHIEN, Y-L. (2013) Which is the better option? Quicker or cheaper? *Psychology & Marketing*, 30, 90-101.
- FELDMAN, L. P., & HORNICK, J. (1981) The use of time: an integrated conceptual model. *The Journal of Consumer Research*, 7, 407-419.
- GRAHAM, R. J. (1981) The role of perception of time in consumer research. *Journal of Consumer Research*, 7, 335-342.
- GROSS, B. L. (1987) Time scarcity: interdisciplinary perspectives and implications for consumer behavior. J. N. Sheth & E. Hirschman (Eds.), *Research in consumer behavior*. Greenwich, UK: JAI. Pp. 1-54.
- KAHNEMAN, D., & TVERSKY, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica*, 47, 263-291.
- LECLERC, F., SCHMITT, B. H., & DUBÉ, L. (1995) Waiting time and decision making: is time like money? *Journal of Consumer Research*, 22, 110-119.
- LINVILLE, P. W., & FISCHER, G. W. (1991) Preferences for separating or combining events. *Journal of Personality and Social Psychology*, 60, 5-23.
- OKADA, E. M., & HOCH, S. J. (2004) Spending time versus spending money. *Journal of Consumer Research*, 31, 313-323.

- SAINI, R., & MONGA, A. (2008) How I decide depends on what I spend: use of heuristics is greater for time than for money. *Journal of Consumer Research*, 34, 914-922.
- SOMAN, D. (2001) The mental accounting of sunk time costs: why time is not like money. *Journal of Behavioral Decision Making*, 14, 169.
- SOSTER, R. L., MONGA, A., & BEARDEN, W. O. (2010) Tracking costs of time and money: how accounting periods affect mental accounting. *Journal of Consumer Research*, 37, 712-721.
- THALER, R. (1985) Mental accounting and consumer choice. *Marketing Science*, 4, 199-214.
- THALER, R. H. (1999) Mental accounting matters. *Journal of Behavioral Decision Making*, 12, 183-206.
- THALER, R. H., & JOHNSON, E. J. (1990) Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. *Management Science*, 36, 643-660.
- ZAUBERMAN, G., & LYNCH, J. G. (2005) Resource slack and propensity to discount delayed investments of time versus money. *Journal of Experimental Psychology: General*, 134, 23-37.

Accepted September 12, 2013.

APPENDIX A

You are going to participate in a great sale. After arriving at the site of the sale, you notice that you will have to wait in line for approximately six hours. Fortunately, the sponsor allows all participants to take a break for any two hours among six hours in line. In other words, you can take a break of two hours from the line, and it will not affect your right to purchase or your position in line.

Although your waiting time in line will decrease from six hours to four hours, your total time expenditure is still six hours (i.e., break time of two hours and waiting time of four hours). Selecting the first hours to take a break cannot allow you to arrive at the sale place later, and choosing the last hours to rest will not permit you to purchase earlier. Please just consider two discretionary hours among the six hours of waiting time to make your decision.

APPENDIX B

Imagine you are a temporary worker on assignment. One day, you receive two different assignments from two different companies. For each assignments, you must work eight hours (from 9 A.M. to 5 P.M.), and you could work any two days next week. Assume that the two assignments pay the same; furthermore, the effort and traffic costs of both assignments are also the same. How will you arrange the two assignments next week (from Monday to Sunday)?

APPENDIX C

This weekend, you have two unplanned tasks and must spend your discretionary time to complete them. Firstly, you have to get concert tickets, and you expect this task will require 30 minutes of waiting. Secondly, you have to go to the bike shop to maintain your motorcycle. You know this will include 15 minutes of waiting. Now, assuming both places are open on Saturday and Sunday, and traffic is at the same level on both days, if you must complete two tasks this weekend would you deal with both on the same day (Saturday or Sunday) or on different days (Saturday and Sunday)? Please write down your reasons.

WHY DO PHYSICIANS TREAT THEIR RELATIVES? EXPLORING THE INFLUENCE OF SOCIAL PSYCHOLOGY^{1, 2}

JONATHAN R. SCARFF

Veterans Affairs Outpatient Clinic

Summary.—Physicians often receive requests for treatment, medical advice, or other intervention from relatives. Most doctors comply. Reasons for compliance can be categorized by doctors' attitudes toward the relative, colleagues, themselves, ethical guidelines, and the problem. Compliance may be influenced by elements of social psychology as well. Social exchange theory, persuasion techniques, attribution, conformity, desire for approval, and the affinity principle can induce intervention. Future research should explore doctors' attitudes toward relatives, the medium by which requests are made, treatment outcomes, changes in the relationship following treatment, cultural or familial norms, and changes in clinicians' beliefs or behavior that occur when facing opposing requests and guidelines. Awareness of these influences may help physicians to make objective decisions regarding intervention.

For centuries, doctors have received requests for medical treatment, advice, or other intervention from relatives, and surveys have found that most physicians comply with such requests. Thomas Percival first addressed potential ethical and practical conflicts and discouraged physicians from treating family members (Percival, 1803). Citing potential for compromised objectivity, incomplete evaluations, unwillingness to discuss sensitive topics, medical errors, or interpersonal problems following a bad outcome, this recommendation is espoused by medical organizations (American Medical Association, 1993; American Academy of Pediatrics, 2009). The most recent code of medical ethics generally discourages intervention except for short-term, minor problems, emergencies, or illnesses in isolated settings (American Medical Association, 1993). These circumstances are not defined, so decisions to treat are influenced by clinicians' attitudes toward multiple entities (Scarff & Lippmann, 2012). Previously unexplored in the literature is the potential explanatory power of social psychology principles and theories. This paper reviews reasons for intervention and explores these principles that may underlie decisions to treat.

¹Address correspondence to Jonathan R. Scarff, Veterans Affairs Outpatient Clinic, 279 North Grove Medical Park Drive, Spartanburg, SC 29303 or e-mail (Jonathan.Scarff2@va.gov).

²The author gratefully acknowledges Mary Ellen Wojtasiewicz, R.N., Ph.D., for her comments regarding the manuscript. The views expressed in this article are those of the author and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

METHOD

A literature search of the PsycINFO and Ovid MEDLINE databases from January 1950 to June 2013 was conducted using the search words "family member," "relatives," "treatment," "prescribing," "physician," and "ethics." Eleven surveys were identified and evaluated to identify reasons that physicians may have given for intervention.

RESULTS

Detailed descriptions of the surveys, their methodology, respondents, relatives, and treatments are described elsewhere (Scarff & Lippmann, 2012). All surveys reported each physicians' specialty, demographic information, the relative receiving care, and interventions offered, but none assessed treatment outcomes. Reasons for intervention can be grouped based on the clinician's attitude toward various entities, which are described later. Only eight surveys solicited reasons for intervention, and there were no measures of validity or reliability performed on survey data except by Chen and colleagues (Chen, Feudtner, Rhodes, & Green, 2001). The reasons given are examined to explore principles of social psychology which may explain decisions to intervene.

A mailed survey of 94 family medicine physicians in South Carolina (78% responded) assessed the frequency with which they treated spouses (compared to a control group of lawyers). A pilot questionnaire had been administered to resident physicians and attorneys to improve clarity and ease of data gathering. Physicians reported treating more head, ear, throat, and back pain, vomiting, and depression (Boiko, Schuman, & Rust, 1984). Another pretested, structured questionnaire was given to 691 primary care and specialist doctors at a teaching hospital in Illinois (67% responded) to assess the frequency with which doctors performed various procedures for relatives. Results indicated that 95% of doctors (most commonly primary care physicians over 45 years of age) provided treatment. Requests ranged from frequently prescribing medications to infrequently performing emergency surgery; spouses, children, mothers, and siblings made requests most frequently (La Puma, Stocking, La Voie, & Darling, 1991). A Likert-scale questionnaire listing types of illnesses and physical or laboratory assessments was mailed to 701 academic generalist and specialty physicians in Iowa to explore how frequently they treated their own children; 70% responded (Dusdieker, Murph, Murph, & Dungy, 1993). Most (74%) of the respondents had treated acute afebrile illnesses, and 65% had prescribed medications (Dusdieker, *et al.*, 1993).

Randomly selected by ZIP code, 2,014 primary care doctors and surgeons (general, orthopedic, and obstetrics/gynecologists) in urban, suburban, and rural areas in Oregon received a mailed survey assessing the

frequency and comfort with which they had provided various interventions; 65% responded (Reagan, Reagan, & Sinclair, 1994). Comfort was assessed using a Likert-type scale. Many clinicians reported prescribing medications, providing pediatric care, and performing physical examinations (Reagan, *et al.*, 1994). Family and internal medicine resident doctors in Delaware received a survey to assess willingness to write prescriptions in 12 hypothetical scenarios (Aboff, Collier, Farber, & Ehrenthal, 2002). Of 92 doctors surveyed, 80% responded, and 95% said they would give antibiotics to a visiting sibling, and 77% would treat their own child for ear infection (Aboff, *et al.*, 2002). Randomly chosen from a Internet directory, 1,086 pediatricians and pediatric subspecialists received an e-mailed and subsequent paper survey assessing treatment practices toward relatives, particularly children (Walter, Pappano, & Ross, 2009). Of 430 respondents, most had prescribed medications or performed physical examinations (Walter, *et al.*, 2009).

Eight academic family medicine physicians throughout the United States with an ill parent received a semi-structured, in-depth telephone interview to assess potential role conflicts. Interviews were transcribed and independently coded for themes using an open-coding technique (Chen, *et al.*, 2001). Participants were re-interviewed to validate and clarify themes from initial interviews. They identified competing personal and professional expectations assigned to them by colleagues and relatives. Five family medicine doctors in Oregon gave unstructured narratives regarding involvement in their spouses' or children's care, ranging from requesting radiology results to expediting hospital care; no data analysis was conducted (Fromme, Farber, Babbott, Pickett, & Beasley, 2008).

Reasons for intervention can be classified by the doctors' attitude toward (or perception of) the relative, colleagues, themselves, guidelines/laws, or the situation. Desiring to reciprocate or compensate for financial or emotional support, feeling obligated to honor a request, "bowing to pressure" or "wanting power" (these reasons listed but not elucidated by Boiko and colleagues), wishing to maintain a good relationship, and avoiding guilty feelings by denying care were reasons for intervention (Boiko, *et al.*, 1984; Reagan, *et al.*, 1994; Aboff, *et al.*, 2002). Desiring convenience, financial savings, and confidentiality for the relative were additional reasons (Boiko, *et al.*, 1984; La Puma, *et al.*, 1991; Dusdieker, *et al.*, 1993; Reagan, *et al.*, 1994; Fromme, *et al.*, 2008; Walter, *et al.*, 2009).

Doctors' attitudes toward colleagues and themselves are additional reasons for treatment. Treatment may occur if a colleague is perceived as untrustworthy, overworked, or unavailable (Dusdieker, *et al.*, 1993; Reagan, *et al.*, 1994; Walter, *et al.*, 2009). Practitioners may intervene if they believe that they possess greater skill, concern, or knowledge compared

to colleagues (La Puma, *et al.*, 1991; Dũsdieker, *et al.*, 1993; Reagan, *et al.*, 1994; Chen, *et al.*, 2001; Walter, *et al.*, 2009).

Attitudes to guidelines, the problem, and the circumstances may motivate intervention. Doctors may intervene if no local law exists, if they are unaware of or choose to disregard guidelines, or if they feel that laws or guidelines do not apply to them (Boiko, *et al.*, 1984). They may intervene if they feel the problem is minor, emergent, or in an isolated setting (Boiko, *et al.*, 1984; La Puma, *et al.*, 1991; Reagan, *et al.*, 1994; Fromme, *et al.*, 2008).

DISCUSSION

Although surveys indicate that intervention is common, no study has explored how principles of social psychology might explain intervention. Potential influential principles include, but are not limited to, social exchange theory (e.g., the reciprocity norm), persuasion techniques, attribution messages, conformity, desire for approval, and the affinity principle.

Social exchange theory is an "exchange of activity, tangible or intangible, and more or less rewarding or costly, between at least two persons" (Homans, 1961). This theory prevents violation of social norms by various and subtle means, including ingratiation. One societal norm is the reciprocity norm: the positive reciprocity norm is a feeling of obligation to repay a favor, and malcontent can follow if the favor goes unpaid; negative reciprocity refers to unfavorable treatment being reciprocated (Gouldner, 1960). In a family study, reciprocity depended on perceived intergenerational support; perceived inequity of support negatively affected the relationship (Schwarz, Trommsdorff, Albert, & Mayer, 2005). Adult children were more likely to engage in reciprocal interactions with parents if parents were dependent, received intense support, and had sufficient opportunities to reciprocate (Leopold & Raab, 2011). The reciprocity norm is alluded to in the Boiko, *et al.* (1984) and Aboff, *et al.* (2002) surveys, where doctors would treat relatives in exchange for earlier financial or emotional support.

Persuasion techniques may elicit compliance, such as the "foot-in-the-door" technique, the "door-in-the-face" technique, the manner of request, and attribution messages. In the "foot-in-the-door" technique, complying with a small request increases the likelihood of complying with a larger request from the same person (Freedman & Fraser, 1966). For example, if a relative has received a physical examination, the doctor may comply with a request for medication as the next step. In the "door-in-the-face" technique, a large initial request is usually denied but is followed by a smaller request which is usually met (Cialdini, Vincent, Lewis, Catalan, Wheeler, & Darby, 1975). For example, a physician may deny a significant request (e.g., prescribe controlled medication), but then comply with a minor request (e.g., complete a sports physical). Requests are more effective when made in person rather than through another medium (Jecker,

Maccoby, Breitrose, & Rose, 1964). Although no survey reviewed the manner in which relatives made requests, face-to-face requests should be met more frequently than those made via another medium (e.g., telephone, another relative). In one study, individuals receiving statements (e.g., "You are tidy") of positive internal attributes were likely to comply with a request (clean a work area) because the statement fit the individual's self-concept; persuasive messages (e.g., "You should be tidy") elicited a negative reaction from being misinterpreted as implying that individuals did not possess the desired attribute (Miller, Brickman, & Bolen, 1975). Therefore, doctors may be more likely to meet requests combined with positive attributions (e.g., "You are helpful and caring") rather than persuasive messages (e.g., "You should be helpful and caring.")

Conformity, need for approval and the affinity principle may explain intervention. Pressure toward conformity exists in groups, and an individual in a group is likely to conform to the attitudes, beliefs, and behaviors of the majority (Festinger, 1950; Latané, 1996). Therefore, physicians surrounded by colleagues who commonly treat relatives may ultimately change their opinions and behavior to become similar to those of the colleagues, even if they are aware of ethical conflicts or averse to the practice. People with a high need for approval tend to be easily persuaded (Vallacher, Nowak, & Miller, 2003). The basic desire to be liked and approved, especially by one's own relatives, may be sufficient reason to treat relatives. Finally, the affinity principle states that people tend to be more compliant with requests made by people they like (Crowne & Marlowe, 1964). Therefore, doctors are expected to comply with requests made by liked family members more than those made by disliked relatives.

Surveys have noted reasons for and against intervention, but have not identified elements of social psychology which may engender compliance with requests. This overview explores some principles of social psychology that may influence treatment decisions. Limitations include inclusion of a small number of English-language surveys written after 1950, with a limited number of respondents. Survey responses may have been distorted by recall or social desirability biases, and most did not test reliability or validity.

Future surveys should explore the medium by which requests are made, physicians' attitudes toward and relationship with relatives, and treatment outcomes, which have been previously unexplored. Doctors have reported positive (Horn, 1990; Klein, 1997) and negative outcomes (Rosenbaum, 1992; Varga, 2007) from intervention, but research using a larger sample can objectively assess treatment outcomes and possible changes in the relationship. If treatment in certain scenarios tends to result in worse outcomes, this would be an additional reason to discourage physicians from

intervention. Advances in technology (e.g., Internet, social networks) can affect family relationship dynamics and should be considered when developing updated ethical guidelines. Other research opportunities include consequences to the relationship if a physician declines intervention, cultural or familial norms that may influence whether physicians are likely to receive or comply with requests, and physicians' attitudes toward treatment and guidelines. Greater knowledge about ethical guidelines, reasons for and against treatment, and basic social psychology principles may enable doctors to make informed decisions and assist relatives in obtaining appropriate medical care.

REFERENCES

- ABOFF, B. M., COLLIER, V. U., FARBER, N. J., & EHRENTAL, D. B. (2002) Residents' prescription writing for nonpatients. *Journal of the American Medical Association*, 288, 381-385.
- AMERICAN ACADEMY OF PEDIATRICS. (2009) From the American Academy of Pediatrics: policy statements—pediatrician-family-patient relationships: managing the boundaries. *Pediatrics*, 124, 1685-1688.
- AMERICAN MEDICAL ASSOCIATION. (1993) Section 8.19: Self-treatment or treatment of immediate family members. *Code of Medical Ethics*. Retrieved June 11, 2013, from <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion819.page>.
- BOIKO, P. E., SCHUMAN, S. H., & RUST, P. F. (1984) Physicians treating their own spouses: relationship of physicians to their own family's health care. *Journal of Family Practice*, 18, 891-896.
- CHEN, F. M., FEUDTNER, C., RHODES, L. A., & GREEN, L. A. (2001) Role conflicts of physicians and their family members: rules but no rulebook. *Western Journal of Medicine*, 175, 236-239.
- CIALDINI, R. B., VINCENT, J. E., LEWIS, S. K., CATALAN, J., WHEELER, D., & DARBY, B. L. (1975) Reciprocal concessions procedure for inducing compliance: the door-in-the-face technique. *Journal of Personality and Social Psychology*, 31, 206-215.
- CROWNE, D. P., & MARLOWE, D. (1964) *The approval motive: studies in evaluative dependence*. New York: Wiley.
- DUSIEKER, L. B., MURPH, J. R., MURPH, W. E., & DUNGY, C. I. (1993) Physicians treating their own children. *American Journal of Diseases of Children*, 147, 146-149.
- FESTINGER, L. (1950) Informal social communication. *Psychological Review*, 57, 271-282.
- FREEDMAN, J. L., & FRASER, S. C. (1966) Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4, 195-202.
- FROMME, E. K., FARBER, N. J., BABBOTT, S. F., PICKETT, M. E., & BEASLEY, B. W. (2008) What do you do when your loved one is ill? The line between physician and family member. *Annals of Internal Medicine*, 149, 825-831.
- GOULDNER, A. W. (1960) The norm of reciprocity: a preliminary statement. *American Sociological Review*, 25, 161-178.
- HOMANS, G. C. (1961) *Social behaviour: its elementary forms*. New York: Harcourt, Brace and World.
- HORN, A. A. (1990) Diagnosis and treatment among relatives. *Canadian Medical Association Journal*, 143, 91.

- JECKER, J., MACCOBY, N., BREITROSE, H. S., & ROSE, E. D. (1964) Teacher accuracy in assessing cognitive visual feedback from students. *Journal of Applied Psychology*, 48, 393-397.
- KLEIN, M. C. (1997) Too close for comfort: a family physician questions whether medical professionals should be excluded from their loved ones' care. *Canadian Medical Association Journal*, 156, 53-55.
- LA PUMA, J., STOCKING, C. B., LA VOIE, D., & DARLING, C. A. (1991) When physicians treat members of their own families: practices in a community hospital. *New England Journal of Medicine*, 325, 1290-1294.
- LATANÉ, B. (1996) Dynamic social impact: the creation of culture by communication. *Journal of Communication*, 46, 13-25.
- LEOPOLD, T., & RAAB, M. (2011) Short-term reciprocity in late parent-child relationships. *Journal of Marriage and Family*, 73, 105-119.
- MILLER, R. L., BRICKMAN, P., & BOLEN, D. (1975) Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, 31, 430-441.
- PERCIVAL, T. (1803) *Medical ethics, or, a code of institutes and precepts, adapted to the professional conduct of physicians and surgeons; to which is added an appendix containing a discourse on hospital duties, and notes and illustrations*. Printed by S. Russell, for J. Johnson & R. Bickerstaff: Manchester, England.
- REAGAN, B., REAGAN, P., & SINCLAIR, A. (1994) 'Common sense and a thick hide.' Physicians providing care to their own family members. *Archives of Family Medicine*, 3, 599-604.
- ROSENBAUM, E. E. (1992) When physicians treat their own families. *New England Journal of Medicine*, 326, 895.
- SCARFF, J. R., & LIPPMANN, S. (2012) When physicians intervene in their relatives' health care. *HEC Forum*, 24, 127-137.
- SCHWARZ, B., TROMMSDORFF, G., ALBERT, I., & MAYER, B. (2005) Adult parent-child relationships: relationship quality, support, and reciprocity. *Applied Psychology: An International Review*, 54, 396-417.
- VALLACHER, R. R., NOWAK, A., & MILLER, M. E. (2003) Social influence and group dynamics. In T. Millon & M. J. Lerner (Eds.), *Comprehensive handbook of psychology*. Vol. 5. *Personality and social psychology*. New York: Wiley. Pp. 383-417.
- VARGA, O. (2007) Physicians do not heal thy relatives. *Patient Education and Counseling*, 67, 11-12.
- WALTER, J. K., PAPPANO, E., & ROSS, L. F. (2009) A descriptive and moral evaluation of providing informal medical care to one's own children. *Journal of Clinical Ethics*, 20, 353-361.

Accepted September 12, 2013.

MEASUREMENT OF ACADEMIC ENTITLEMENT¹

BRIAN K. MILLER

Texas State University

Summary.—Members of Generation Y, or Millennials, have been accused of being lazy, whiny, pampered, and entitled, particularly in the college classroom. Using an equity theory framework, eight items from a measure of work entitlement were adapted to measure academic entitlement in a university setting in three independent samples. In Study 1 ($n = 229$), confirmatory factor analyses indicated good model fit to a unidimensional structure for the data. In Study 2 ($n = 200$), the questionnaire predicted unique variance in university satisfaction beyond two more general measures of dispositional entitlement. In Study 3 ($n = 161$), the measure predicted unique variance in perceptions of grade fairness beyond that which was predicted by another measure of academic entitlement. This analysis provides evidence of discriminant, convergent, incremental, concurrent criterion-related, and construct validity for the Academic Equity Preference Questionnaire.

The positive relation between perceptions of the fairness of rewards and satisfaction with rewards is well documented and straightforward (Cohen-Charash & Spector, 2001; Colquitt, Conlon, Wesson, Porter, & Ng, 2001) for most individuals. However, persons who perceive themselves to be entitled consistently judge that they are rewarded unfairly and they are often dissatisfied with their outcomes in certain exchange relationships. Entitlement is a trait on the equity sensitivity spectrum, with benevolence at the opposite pole and equity sensitivity in the middle range (Huseman, Hatfield, & Miles, 1985, 1987), whereby persons compare what they get from a situation (outcome) to what they bring to the situation (input) to determine if they are rewarded fairly (Adams, 1963, 1965). In particular, today's youth (variously referred to as Generation Y or Millennials) insist more often than previous generations that they are rewarded unfairly regardless of their effort or contribution, and they are sometimes perceived by employers, co-workers, and teachers as more entitled than previous generations (Stout, 2000; Twenge, 2006; Twenge, Konrath, Foster, Campbell, & Bushman, 2008a, 2008b; Twenge, Campbell, Hoffman, & Lance, 2010).

Stout (2000) suggested that the lowering of educational standards has given rise to an entire generation of young persons who demand high grades for low performance and who are unhappy when their expectations regarding grades are unmet. To adequately capture this emerging

¹Address correspondence to Brian K. Miller, Texas State University, Department of Management, 545 McCoy Hall, San Marcos, TX, USA, 78666 or e-mail (bkmiller@txstate.edu).

construct, the search for a psychometrically reliable and valid domain-specific measure of academic entitlement has begun in earnest (e.g., Achacoso, 2002; Greenberger, Lessard, Chen, & Farruggia, 2008; Chowning & Campbell, 2009; Kopp, Zinn, Finney, & Jurich, 2011). The purpose of this paper was to adapt a measure of work entitlement to another specific domain to measure academic entitlement for college students. Domain-specificity in the measurement of entitlement is important, because students who display academic entitlement may not display such tendencies in other domains or with non-academic others (Chowning & Campbell, 2009; Kopp, *et al.*, 2011) like family, health care providers, and friends outside their college experience.

The need for a domain-specific measure of a stable individual difference in personality motivates this study as it has others. The measure of academic entitlement by Chowning and Campbell (2009) was developed largely as a predictor of student incivility. Entitlement as a dispositional predictor of uncivil behavior can be traced back to Freud's psychoanalytic structural model of the id, ego, and superego (Freud, 1961) or object relations theory (Moore, 1975; Sandler, Holder, & Dare, 1976; Tiecholz, 1978; Fine, 1986). Kopp, *et al.*'s (2011) items were based on theoretical underpinnings of other entitlement researchers and essentially took the best of most previous efforts. Entitlement, in general, explains differing reactions to unfairness; an academic entitlement scale might best be developed out of a theory of reactions to (un)fairness, like equity theory.

The theoretical framework for the current study can be traced back to Festinger's social comparison theory (1954), his theory of cognitive dissonance (1957), Adams' equity theory (1963, 1965), equity sensitivity (Huseman, *et al.*, 1985, 1987), and to the Equity Preference Questionnaire (Sauley & Bedeian, 2000) adapted here. At the heart of this sequence is equity theory, which describes the experience of dissonance when one's outcome-to-input ratio is not equivalent to the ratio of some comparison person. Thus, dissonance is a drive-like emotion that must be ameliorated by the resolution of the inequity experienced from the comparison. Entitled persons prefer more positive outcomes in relation to their inputs as compared to everyone else (Huseman, *et al.*, 1985, 1987). They are never satisfied with their rewards unless they are comparatively over-rewarded. Well-documented outcomes of equity theory include reward fairness and pay satisfaction.

For college students, rewards often come in the form of grades. Thus, it is hypothesized that among college students, academic entitlement will be negatively related to both their university satisfaction and their perception of grade fairness. For example, an entitled student may insist that he was graded unfairly simply because he was dissatisfied with the grade he

earned. In this study, academic entitlement is generally defined as preferring to receive more from one's academic experience than one's peers and preferring to get more from one's academic experience than one gives to it. More specifically, academic entitlement is the perception that an individual is entitled to higher grades than she earned or than other students earned regardless of how hard the individual studied, the amount of time she spent in preparation, or her own ability.

Etiology of Entitlement

Although the etiology of an inflated concern for self so inherently important to broadly measured dispositional entitlement involves considerable debate (Millon, 1981), post-World War II generations have often complained about the attitudes and behaviors of members of the generations that followed them. Mosak (1959) attributed the post-World War II rise in entitlement in that generation to overly permissive child-rearing practices. Later, Lasch (1978) suggested that protests of the Vietnam War fostered disillusionment that led to unusually high self-centeredness, materialistic consumption, and the lack of capacity for sustained intimacy for some members of that generation (but certainly not all). Nevertheless, all of these are elements of entitlement. Huseman, *et al.* (1987) suggested that pampered children grow to become entitled adults who are more predisposed toward dissatisfaction except when they are receiving what objectively might be considered to be unearned rewards. This inflated sense of self-importance and uniqueness has given rise to a preoccupation with receiving attention and an entitlement to special treatment by others (Bergman, Westerman, & Daly, 2010). In efforts to place entitlement in the nomological network, some researchers (e.g., Stout, 2000; Twenge, 2006; Trzesniewski, Donnellan, & Robins, 2008; Twenge, *et al.*, 2008a) have focused on dispositional entitlement as a component of narcissism or as a stand-alone disposition (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004).

Researchers have suggested that today's college-aged persons are more narcissistic than in years past, although research on intergenerational differences in scores on the Narcissistic Personality Inventory (see Raskin & Hall, 1979; Raskin & Terry, 1988) range from moderate, with a shift in scores of around 1/3 standard deviation over 25 years in over 85 samples collected from similarly aged cohorts (Twenge, *et al.*, 2008a), to no change (Trzesniewski, *et al.*, 2008). In general, the lay perception of Generation Y is that it is more egotistical, overconfident, and entitled than previous generations of young persons at the same age (Twenge, 2006). Such a sense of entitlement likely grew, for some, out of the self-esteem movement of the 1990s (Twenge, 2006). Spearheading this movement, Mecca, Smelser, and Vasconcellos (1989) suggested that high self-esteem would cure all that ailed the younger generation. Additionally, they asserted that high self-esteem

would lead to high performance, and not vice versa as was commonly thought.

Such efforts at increasing self-esteem (e.g., Branden, 1987, 1994) have led to a generation of young persons who have learned that it is normal to expect automatic admiration and praise and that their behaviors should be valued regardless of actual accomplishments (Millon, 1998; Millon & Davis, 2000). The self-esteem movement (Mecca, *et al.*, 1989) has been characterized as misplaced and unfounded (Stout, 2000), poorly conceived (Twenge, 2006), and of little value (Baumeister, Campbell, Krueger, & Vohs, 2005). The emphasis on self-esteem at all costs, including the reduction of academic rigor in curricula (Stout, 2000), has resulted in "widespread concern that the American 'feel-good culture' and the self-esteem programs it spawned have inadvertently produced a generation of young narcissists" (Trzesniewski, *et al.*, 2008).

However, the relationship between narcissism and self-esteem is not clear-cut (Ziegler-Hill, Myers, & Clark, 2010) and both likely involve more than just being self-centered (Lasch, 1978); it appears that only fragile self-esteem rooted in instability is related to narcissism (Rhodewalt, Madrian, & Cheney, 1998; Ziegler-Hill, Chadha, & Osterman, 2008). It is true that self-esteem can yield many benefits related to mental health and motivation (Ryan & Brown, 2006), but self-esteem that is mismatched with actual accomplishments can lead to insensitivity, self-centeredness, and poor regard for others, each of which are aspects of entitlement (Paulhus & Williams, 2002). Bogart, Benotsch, and Pavlovic (2004) suggest that narcissism can also be affected by comparisons with others and that certain components of narcissism have different relationships to different social comparisons. Specifically, Bogart, *et al.* found that persons scoring high on the entitlement component of narcissism feel validated when their worldview matches their experiences, but they often experience anxiety and anger when others get more than they do. Mosak (1971) suggests that entitled persons tend "to view life as unfair for denying him that to which he is entitled" (p. 78). This sense of entitlement is derived from social comparisons (Festinger, 1954) and the discomfort that the resulting dissonance (Festinger, 1957) creates is strongly tied to the concept of equity sensitivity (Huseman, *et al.*, 1985; Sauley & Bedeian, 2000).

Workplace Entitlement

Equity sensitivity was originally conceived of as an attitudinal construct (Huseman, *et al.*, 1985) and later as an individual difference; people have differing sensitivity (Huseman, *et al.*, 1987), tolerance (King, Jr., Miles, & Day, 1993), or preference (Sauley & Bedeian, 2000) for unfairness in their workplace exchange relationships. Sauley and Bedeian (2000) developed the

Equity Preference Questionnaire to overcome some measurement shortcomings of the Equity Sensitivity Instrument (Huseman, *et al.*, 1985), which uses a forced-choice response system. Sauley and Bedeian conceived of equity sensitivity as unidimensional and created 16 items to measure opposite ends of the equity sensitivity continuum (i.e., benevolence and entitlement as described by King, *et al.*, 1993), with half the items requiring reverse scoring.

To date, two published examinations of the dimensionality of the Equity Preference Questionnaire exist (e.g., Foote & Harmon, 2006; Miller, 2009). Using principal components analysis, Foote and Harmon (2006) found that each reverse-scored item designed to measure entitlement loaded on the same single component and the other items designed to measure benevolence were split between two additional components. These latter two components were correlated with the entitlement component at $-.56$ ($p < .01$) and $-.34$ ($p < .01$). Accordingly, these two benevolent components were positively correlated with each other at $.33$ ($p < .01$). Foote and Harmon suggested “that the (Equity Preference Questionnaire) may not be a unidimensional scale” (p. 98), that the items in the scale do not measure the same thing, and “at the very least, one should not treat this scale as unidimensional” (p. 103). Extending Foote and Harmon’s research, Miller (2009) examined the factor structure of the 16-item Equity Preference Questionnaire using confirmatory factor analysis (CFA) and found that when forcing the eight ‘Entitlement’ items and the eight ‘Benevolent’ items to load on separate factors (i.e., not allowing cross-loading), a statistically significant improvement to fit was found over a one-factor structure comprised of all 16 items. With this research in mind, both Foote and Harmon (2006) and Miller (2009) urged subsequent researchers to examine the Benevolent items and the Entitlement items separately.

Academic Entitlement

Equity sensitivity, and its components of Entitlement and Benevolence, is an individual difference whose measurement has principally been of concern to organizational researchers. Educational researchers have recently developed measures of entitlement for a different domain: coursework experiences. This requires the development of a domain-specific measure of academic entitlement (Greenberger, *et al.*, 2008; Kopp, *et al.*, 2011), rather than relying on the entitlement subscale of the Narcissistic Personality Inventory (Raskin & Terry, 1988) or a more general dispositional measure of entitlement like the Psychological Entitlement Scale (Campbell, *et al.*, 2004). To address this, Chowning and Campbell (2009) conducted both principal components analysis (with an orthogonal rotation suggesting that the resulting factors are unrelated) in the item-reduction phase of their scale development and confirmatory factor analysis (CFA) on the final version of their mea-

sure of academic entitlement. They found that a two-factor structure fit the data best and described the factors of academic entitlement as externalized responsibility and entitled expectations. Items measuring the former focused on the relationship between effort and reward on group projects, disdain for teachers in general, and relying on others. The other subscale focused on the expectations of grades as well as the obligation of teachers. Of note is that for the three samples for which Chowning and Campbell reported alpha reliabilities, scores on the Entitled Expectations subscale each failed to reach .70, a cutoff largely considered minimal for exploratory non-clinical research (Nunnally, 1978). The Externalized Responsibility and Entitled Expectations factors were correlated at only $r = .21$ ($p < .001$); thus, Chowning and Campbell (2009) recommended that researchers treat these two factors as distinct and not sum them for a total score. The Externalized Responsibility subscale correlated with the Psychological Entitlement Scale at $r = .38$ ($p < .01$) in one sample and $r = .28$ ($p < .01$) in another sample and with the Narcissistic Personality Inventory at $r = .18$ ($p < .01$) in the first sample and at $r = .09$ ($p < .05$) in the second sample. Scores on the Entitled Expectations subscale correlated with the Psychological Entitlement Scale at $r = .18$ ($p < .01$) and $r = .34$ ($p < .01$) in two samples but weakly with the Narcissistic Personality Inventory ($r = .10$, ns, in the first sample, $r = .18$, $p < .01$, in the second sample). These independent subscale scores were also used to predict reactions to negative feedback in a laboratory using stepwise regression, but as Thompson (1995) suggests, the use of stepwise regression is never appropriate because it tends to capitalize on the chance characteristics of the sample.

Current Study

The items from the entitlement portion of the Equity Preference Questionnaire (hereafter referred to as the Academic Equity Preference Questionnaire, AEPQ) were adapted to a university setting by replacing mention of a work organization or employer with a reference to the respondents' coursework at a university. Data were collected at a mid-sized public university in the southwestern United States in three independent samples gathered in large sections of a junior-level business course required of both business majors and business minors, but closed to all others. All data were collected via anonymous self-report paper-and-pencil surveys in exchange for extra credit. In Study 1 ($n = 229$), confirmatory factor analysis (CFA) was used to examine scores on the new AEPQ. In Study 2 ($n = 200$), the AEPQ was compared to more general dispositional measures of entitlement in the prediction of university satisfaction. In Study 3 ($n = 161$), the AEPQ was compared to Chowning and Campbell's (2009) academic entitlement measure in the prediction of grade fairness.

Study 1: Confirmatory Factor Analysis of the Adapted Scale
METHOD

Participants

Data were collected via a confidential survey from 229 respondents enrolled in one of two sections of an introductory Management course required of both business majors ($n=150$) and business minors, with fields of major study ranging from fashion merchandising to manufacturing engineering. Student t tests for differences between groups of business majors and minors revealed no significant differences in scores on the AEPQ. The average age of the sample was 22.6 yr., 137 respondents were male, and the self-reported racial/ethnic group membership was as follows: 76.9% White, 3.5% African-American, 14.1% Hispanic, 3.0% Asian, and 2.6% Other. Correlations with age, Student t tests for differences based on sex, and chi-squared tests for differences among racial or ethnic groups were not significant on the AEPQ.

Academic Equity Preference Questionnaire

Academic entitlement was measured by adapting, for a university setting, the eight items measuring entitlement from the Equity Preference Questionnaire (Sauley & Bedeian, 2000). The adapted items were: (1) "I prefer to do as little as possible in my classes while getting as much as I can from them," (2) "I am most satisfied in my classes when I have to do as little as possible," (3) "When I am in class, I think of ways to get out of my assignments," (4) "If I could get away with it, I would try to work just a little bit slower in my classes than is expected of me," (5) "It is really satisfying to me when I can get something for nothing in my classes," (6) "It is the smart student who gets as much as they can while giving as little as possible in return," (7) "Students who are more concerned about what they can get in their classes rather than what they can give in their classes are the wise ones," and (8) "If I had to work hard all day for my classes, I would probably quit." Statements were rated on a five-point Likert response scale anchored by 1: Strongly disagree and 5: Strongly agree. Cronbach's coefficient α of internal consistency reliability for scores on the AEPQ items in this sample was .78. Additionally, test-retest reliability analysis was conducted. Six weeks after the completion of the survey, 102 of the initial respondents in one of the sections of the course completed the AEPQ again. Ninety-six provided complete data. The correlation between administrations of the scale at Time 1 and Time 2 six weeks later was .65 ($p < .001$).

Confirmatory Factor Analysis

Confirmatory factor analysis was used to examine the model fit and factor loadings of the items on a single underlying dimension. Because

the chi-squared test is greatly affected by sample size and because CFA is a large sample technique, several supplementary goodness-of-fit indices were used to assess model fit. The standardized root mean square residual (SRMR) ranges from 0 to 1, and values less than .08 indicate good fit (Hu & Bentler, 1999). The root mean square error of approximation (RMSEA: Browne & Cudeck, 1993) ranges from 0 to 1, and values less than .08 indicate reasonable fit, while values less than .05 indicate good fit (Hu & Bentler, 1999). The comparative fit index (CFI: Bentler, 1990) ranges from 0 to 1, with values greater than or equal to .95 indicating good fit (Hu & Bentler, 1999).

Item-total correlations ranged from $r = .23$ to $.53$ ($p < .001$). Item-level skewness values ranged from 0.00 to 1.15 and item-level kurtosis values were from $-.91$ to 2.88 , each of which is below the cutoffs described by West, Finch and Curran (1995) of $|2.0|$ for skewness and $|7.0|$ for kurtosis. Univariate normality at the item level is a necessary but not sufficient condition for multivariate normality (Henson, 1999). Therefore, the normalized Mardia's kurtosis of multivariate normality score of 5.71 was calculated using the DeCarlo (1997) macro. This value is greater than the cutoff of $|3.0|$ recommended by Bentler (1998) and Bentler and Wu (2002). With these statistics in mind the data were univariate but not multivariate normal, so the Satorra-Bentler scaled chi-square ($SB \chi^2$) and robust standard errors adjustment to the maximum likelihood method of estimation was used in CFA with Lisrel 8.80 software (Jöreskog & Sörbom, 2006). This required the computation and input of both the covariance matrix and the asymptotic covariance matrix. The error terms were not allowed to correlate, and the variance of the latent factor was constrained to 1 to set the metric for the latent construct. See Table 1 for the item level data for the AEPQ.

The CFA results for the eight items forced to load on one factor resulted in $SB \chi^2 = 48.42$ ($df = 20$, $p < .001$). The chi-square per degree of freedom was 2.42. The RMSEA was .079, indicating reasonable fit, and the CFI and SRMR were 0.958 and .053, respectively, both of which indicated good fit. The standardized factor loadings on Items 1 through 8 of the AEPQ were, respectively, .62, .58, .67, .54, .58, .55, .52, and .51 (all $ps < .001$). Localized misfit was examined using standardized residuals with only a single excessively large (i.e., $>|3.0|$) standardized residual of 4.99 between Items 1 and 2, suggesting an area of minor model misfit. In Study 1, evidence of construct validity was gathered via good model fit, reasonably strong factor loadings, and a sound theoretical framework from a body of literature developed over 25 years of research on equity sensitivity. Study 2, below, examined the discriminant validity, concurrent criterion-related validity, and the incremental validity of the AEPQ in relation to two more

TABLE 1
STUDY 1 ACADEMIC EQUITY PREFERENCE QUESTIONNAIRE ITEM LEVEL STATISTICS AND CORRELATIONS
WITH 95% CONFIDENCE INTERVALS

Item	1	2	3	4	5	6	7	8
2	.53 (.43, .62)							
3	.38 (.26, .49)	.35 (.23, .46)						
4	.29 (.17, .40)	.26 (.13, .38)	.48 (.37, .57)					
5	.36 (.24, .47)	.32 (.20, .43)	.37 (.25, .48)	.29 (.17, .40)				
6	.33 (.21, .44)	.29 (.17, .40)	.33 (.21, .44)	.24 (.11, .36)	.36 (.24, .47)			
7	.23 (.10, .35)	.25 (.12, .37)	.35 (.23, .46)	.26 (.13, .38)	.30 (.18, .41)	.44 (.33, .54)		
8	.30 (.18, .41)	.28 (.16, .40)	.34 (.22, .45)	.30 (.18, .41)	.31 (.19, .42)	.24 (.11, .36)	.30 (.18, .41)	
M	2.66	2.65	1.72	1.93	2.55	2.08	2.45	2.20
SD	1.05	0.96	0.56	0.71	1.07	0.75	0.72	0.61
Skewness	.24	.36	.01	.56	.32	.69	.00	1.15
Kurtosis	-.91	-.61	-.51	.51	-.73	1.37	-.26	2.88

Note.—All correlations significant at $p < .001$.

dispositional measures of entitlement in the prediction of university satisfaction in an independent sample.

Study 2: Validity in Prediction of Satisfaction

It was hypothesized that the AEPQ, the Psychological Entitlement Scale, and the Entitlement subscale of the Narcissistic Personality Inventory all would be negatively related to university satisfaction but positively related to each other, and that the AEPQ would be more strongly related to university satisfaction, in keeping with the specificity matching principle long favored by trait and attitude researchers (Epstein, 1979; Fleeson, 2004).

METHOD

Participants

In an independent sample, data were collected from 200 respondents. Of the respondents, most ($n = 143$; 71.5%) were business majors and the rest were business minors. Student t tests for differences between business majors and minors revealed no statistically significant differences on any of the variables described below. The average age of this sample was 21.8 yr., 86 respondents were male, and the self-reported racial/ethnic group

membership was as follows: 73.5% White, 3.5% African-American, 22.0% Hispanic, 3.5% Asian, 1% American Indian, and 3.5% Other. Correlations with age, Student *t* tests for differences based on sex, and chi-squared tests for differences among racial or ethnic groups were all non-significant on any of the measured variables described below.

Measures

Academic Equity Preference Questionnaire.—The same eight-item scale with the same Likert response options used in Study 1 was used to measure academic entitlement in Study 2. Cronbach's α for scores on this measure in this independent sample was .74.

Psychological entitlement.—The nine-item scale from Campbell, *et al.*'s (2004) measure of psychological entitlement was used for this broad dispositional measure of entitlement. Responses were collected on items like "I honestly feel like I'm just more deserving than others" and "I demand the best because I'm worth it" using a seven-point Likert-type response scale anchored by 1: Strong disagreement and 5: Strong agreement. Cronbach's α for scores on this measure in this sample was .85. Campbell, *et al.*'s (2004) psychological entitlement scale has been shown by Pryor, Miller, and Gaughan (2008) to be correlated .46 ($p < .01$) with the Entitlement subscale of the Narcissistic Personality Inventory (Raskin & Hall, 1979; Raskin & Terry, 1988) described below.

Narcissistic entitlement.—The six-item entitlement subscale of the Narcissistic Personality Inventory (Raskin & Terry, 1988) was used to measure this disposition. Example items include: "I will never be satisfied until I get all that I deserve" and "If I ruled the world it would be a much better place." Respondents used a seven-point Likert-type scale anchored by 1: Strongly disagree and 7: Strongly agree. The entitlement subscale of the Narcissistic Personality Inventory correlates with the other subscales of the Narcissistic Personality Inventory at values ranging from .14 for vanity to .34 for both authority and exhibitionism (Raskin & Terry, 1988). Cronbach's α in this sample for scores on this measure was .77.

University satisfaction.—Satisfaction with the university was measured by using Miller and Nicols' (2011) adaption of the five items from Hackman and Oldham's (1975) satisfaction scale. An example item is: "Generally speaking, I am very satisfied with this university." Respondents used a five-point Likert-type response scale anchored by 1: Strongly disagree and 5: Strongly agree. Miller and Nicols (2011) reported Cronbach's α for scores on the university satisfaction scale of .70 and correlations with the Core Self-Evaluation Scale (Judge, Erez, Bono, & Thoresen, 2003) of .21 ($p < .01$) and with Miller and Nicols' (2011) measure of grade fairness at .22 ($p < .01$). The alpha coefficient of reliability for scores on this scale in the current sample was .74.

Correlation Analysis

Pearson product moment correlations were computed among variables. As hypothesized, the three measures of entitlement were all positively correlated with each other. However, the Psychological Entitlement Scale was correlated with university satisfaction in a direction opposite of that hypothesized, and narcissistic entitlement was unrelated to university satisfaction. Only the AEPQ was correlated with university satisfaction in the hypothesized direction (Table 2).

TABLE 2
STUDY 2 MEANS, STANDARD DEVIATIONS, CORRELATIONS, AND ALPHA RELIABILITIES

	<i>M</i>	<i>SD</i>	1	2	3	4
1. Psychological entitlement	3.12	0.97	.86			
2. Narcissistic entitlement	4.40	1.09	.59‡	.76		
3. AEPQ	2.42	0.59	.23†	.17*	.75	
4. University satisfaction	4.16	0.49	.16*	.11	-.24†	.74

Note.—Cronbach's α internal consistency reliabilities in boldface on diagonal ($n = 200$).

* $p < .05$, † $p < .01$, ‡ $p < .001$.

Regression Analysis

The examination of bivariate correlations does not allow for the statistical control of other variables. It was hypothesized that the more domain-specific measure of a particular form of entitlement, as measured by the AEPQ, would be more strongly negatively related to the domain-specific criterion of university satisfaction than would the two less domain-specific and more dispositional forms of entitlement (psychological entitlement and narcissistic entitlement). This required statistically controlling for the effects of each predictor and therefore multiple regression analysis was used. University satisfaction was regressed on the AEPQ, the Psychological Entitlement Scale, and narcissistic entitlement ($R^2 = .11$). Standardized regression coefficients were significant for the Psychological Entitlement Scale, but in the opposite direction than expected, and for the AEPQ in the expected direction; the coefficient for narcissistic entitlement was not statistically significant (Table 3).

In Study 2, only the AEPQ was related to university satisfaction in the direction hypothesized, and the AEPQ accounted for unique variance in the criterion above and beyond the two less domain-specific forms of entitlement. Evidence of discriminant validity was gathered via statistically significant correlations of an appropriate magnitude, and thus it was far less than the cutoff for collinearity of $|.85|$ described by Kline (2005) between the AEPQ, the Psychological Entitlement Scale, and narcissistic

TABLE 3
STUDY 2 REGRESSION TESTS WITH UNIVERSITY SATISFACTION AS THE CRITERION

Variable	University Satisfaction as Criterion				Effect Size
	Unstandardized Coefficients			Standardized Coefficient	
	B	SE	95%CI	β	
Constant	4.35	.18	4.00, 4.71		
Psychological entitlement	.10	.04	.01, .18	.19*	.02
Narcissistic entitlement	.02	.04	-.05, .10	.05	.00
AEPQ	-.25	.06	-.36, -.13	-.30‡	.08
$F_{3,196} = 7.83‡, R^2 = .11, \text{Adj } R^2 = .09$					

Note.—AEPQ = Academic Equity Preference Questionnaire. Effect sizes are squared semi-partial correlations. **p* < .05, ‡*p* < .001.

entitlement. These results suggest that scores on the AEPQ measure something different from more general dispositional entitlement. Evidence of concurrent criterion-related and incremental validity was gathered with regression results showing that the AEPQ predicted significant unique variance in university satisfaction in the presence of measures of more dispositional entitlement.

In Study 3, below, convergent validity was examined, comparing scores on the AEPQ to another measure of academic entitlement, and more evidence of concurrent criterion-related and incremental validity evidence was gathered.

Study 3: Validity in Prediction of Fairness

It was hypothesized that scores on the different measures of academic entitlement would be positively correlated with each other, providing evidence of convergent validity. Additionally, academic entitlement as measured by the AEPQ (and derived from equity sensitivity which explains differing preferences for fairness) will likely better predict perceptions of fairness than a measure of academic entitlement developed with a different purpose and with a different theoretical underpinning. Thus, it was hypothesized that the AEPQ would better predict grade fairness than Chowning and Campbell's (2009) measure of academic entitlement.

METHOD

Participants

In this third independent sample, data were collected from 161 respondents. Of the respondents, 111 (68.9%) were business majors and the rest were business minors. There were no statistically significant differences between business majors and minors on any of the variables described below. The average age of the sample was 22.6 yr.; 93 (56.7%) were men, 60

were women (36.6%), and eight failed to indicate their sex (6.7%). The self-reported racial/ethnic group membership was as follows: 73.5% White, 3.1% African-American, 20.4% Hispanic, 1.9% Asian, and 1.2% Other. Correlations with age, Student *t* tests for differences based on sex, and chi-squared tests for differences among racial or ethnic groups were all non-significant on any measured variable described below.

Measures

Academic Equity Preference Questionnaire.—The same AEPQ measure as used in Studies 1 and 2 above was used in Study 3. Cronbach's α for scores on the AEPQ in this sample was .73.

Chowning and Campbell's (2009) measure of academic entitlement.—Chowning and Campbell (2009) developed a two-factor measure of academic entitlement with 10 items designed to measure externalized responsibility and five items to measure entitled expectations. Consistent with their recommendation the two subscales were used as stand-alone measures. Data were gathered using a seven-point Likert-type response scale anchored by 1: Strongly disagree and 7: Strongly agree. Examples of items on the Externalized Responsibility subscale include: "If I do poorly in a course and I could not make my professor's office hours, the fault lies with my professor" and "For group work, I should receive the same grade as the other group members regardless of my level of effort." Examples of the items in the entitled expectations subscale include: "My professors should reconsider my grade if I am close to the grade I want" and "I should never receive a zero on an assignment that I turned in." In this sample, Cronbach's α for the Externalized Responsibility subscale was .73 and on the Entitled Expectations subscale was .72.

Grade fairness.—The four-item measure of distributive justice (i.e., the fairness of rewards) developed by Colquitt (2001) and adapted by Miller and Nicols (2011) was used to measure perceptions of grade fairness. Because grades are often seen as a tangible payoff by students (Covington & Mueller, 2001), the fairness of grades are a form of distributive justice. Colquitt (2001) specifically recommends that researchers adapt his distributive justice items for context and appropriateness. In this study, mention of an employer or pay was replaced with mention of college or grades. Examples of items include: (1) "To what extent do your grades reflect the effort that you have put into your classes in college?" and (2) "To what extent are your grades in your classes in college justified, given your performance?" Respondents used a 5-point Likert-type response scale anchored by 1: To a small extent and 5: To a large extent.

Miller and Nicols (2011) reported Cronbach's α for grade fairness scores of .91, .94, and .89 in three independent samples. They also reported correlations between grade fairness with grade point average of .42 ($p < .001$) and

with university satisfaction of .21 ($p < .01$) in one sample. In another sample, Miller and Nicols (2011) found grade fairness to be correlated with these same variables at .37 ($p < .001$) and .29 ($p < .001$), respectively. In a third independent sample, they reported a grade fairness to be correlated with core self-evaluations at .26 ($p < .001$) and with grade point average at .44 ($p < .001$). In the current sample, Cronbach's α for scores on this scale was .87.

Correlation and Regression Results

Pearson correlations were computed for the four measured variables above. All were statistically significant. The AEPQ was more strongly related with Externalized Responsibility than it was with Entitled Expectations. Each of these three measures of academic entitlement was correlated with grade fairness ranging from $-.25$ to $-.33$ (Table 4). These correlations provide some evidence regarding the convergent validity of the AEPQ in relation to the other measure of academic entitlement. Multiple regression was used to gather additional evidence of concurrent criterion-related and incremental validity in the prediction of grade fairness by each of these variables.

TABLE 4
STUDY 3 MEANS, STANDARD DEVIATIONS, CORRELATIONS, AND ALPHA RELIABILITIES

Variable	M	SD	1	2	3	4
1. AE Entitled Expectations	4.50	1.12	.72			
2. AE Externalized Responsibility	2.05	0.69	.38‡	.73		
3. AEPQ	2.42	0.56	.30‡	.57‡	.73	
4. Grade fairness	3.73	0.74	$-.25^{\dagger}$	$-.32^{\dagger}$	$-.33^{\dagger}$.87

Note.—Cronbach's α internal consistency reliabilities in bold on diagonal ($n = 161$). AE = academic entitlement (Chowning & Campbell, 2009), AEPQ = Academic Equity Preference Questionnaire. $^{\dagger}p < .01$, $^{\ddagger}p < .001$.

In a multiple regression with grade fairness as the criterion, R^2 was .15. The beta weights for Entitled Expectations and Externalized Responsibility subscales were not statistically significant, but beta for the AEPQ was ($\beta = -.21$) (Table 5). These results provide evidence of concurrent criterion-related validity for the AEPQ, as well as incremental validity evidence for the AEPQ above and beyond the predictive capacity of Chowning and Campbell's (2009) measure of academic entitlement.

GENERAL DISCUSSION

In this study, using CFA, it was shown that the AEPQ scores can be considered unidimensional, with acceptable internal consistency reliability

TABLE 5
STUDY 3 REGRESSION TESTS WITH GRADE FAIRNESS AS THE CRITERION

Variable	Grade Fairness as Criterion				Effect Size
	Unstandardized Coefficients			Standardized Coefficient	
	B	SE	95%CI	β	
Constant	5.11	.29	4.54, 5.68		
AE Entitled Expectations	-.08	.05	-.19, .02	-.12	.01
AE Externalized Responsibility	-.17	.10	-.37, .03	-.16	.01
AEPQ	-.28	.12	-.51, -.04	-.21*	.03
$F_{3, 157} = 9.17\ddagger$, $R^2 = .15$, Adj $R^2 = .13$					

Note.—AE = academic entitlement (Chowning & Campbell, 2009), AEPQ = Academic Equity Preference Questionnaire. Effect sizes are squared semi-partial correlations. * $p < .05$, $\ddagger p < .001$.

($\alpha > .70$; Nunnally, 1978) and test-retest reliability (.65). The domain-specific AEPQ was the only significant predictor of domain-specific satisfaction (with the university) when compared to two more general measures of dispositional entitlement. Additionally, the AEPQ outperformed another scale designed to measure academic entitlement in the prediction of perception of grade fairness. The domain-specificity of the AEPQ likely led to its ability to predict satisfaction and fairness, as domain-specific outcomes associated with equity theory.

In Study 1, the unidimensional structure of the AEPQ potentially bodes well for further usage given that scales measuring one and only one construct tend to relate in expected and similar manners with other variables. These results are enhanced by the acceptable internal consistency reliability, test-retest reliability, and standardized factor loadings of the scale. Consistent predictive relationships with key outcomes are the goal of psychometric research, and Studies 2 and 3 provide some evidence of that.

In Study 2, two broader, more dispositional measures of entitlement were either unrelated to the domain-specific outcome of university satisfaction or in a direction opposite of that which was hypothesized. The specificity matching principle suggested as much. Students with higher dispositional entitlement, as measured by the Psychological Entitlement Scale, showed higher university satisfaction; only the AEPQ was negatively related to university satisfaction. Of note is that there was a suppression effect in Study 2, as the bivariate correlations between university satisfaction and both the Psychological Entitlement Scale and the AEPQ were smaller than their standardized regression coefficients. Suppressor variables reduce bivariate correlations unless the suppressors are also included in the model. Because three variables served as predictors, three separate regressions for different

pairs of predictors were run one at a time. In a regression with psychological entitlement and narcissistic entitlement, no suppression occurred. In a second regression with psychological entitlement and the AEPQ and in a third regression with narcissistic entitlement and the AEPQ, suppression occurred. Either the AEPQ serves as the suppressor and is critical to understanding the true nature of the relationship between dispositional entitlement and domain-specific satisfaction, or dispositional entitlement (measured in this study with both the entitlement subscale of the Narcissistic Personality Inventory and with the Psychological Entitlement Scale) suppresses the relationship between academic entitlement and university satisfaction. Either way, the relationships between non-domain-specific measures of entitlement in the prediction of the domain-specific measure of university satisfaction are suppressed unless they are also measured with the domain-specific measure of entitlement. This suggests that because of the non-orthogonal nature of dispositional entitlement and academic entitlement, the former must likely be present in some quantity to give rise to the latter. That is, before academic entitlement can be manifested in a meaningful manner a certain level of general entitlement must be present. While the two are not collinear, it is likely rare that academic entitlement could be predictive of meaningful outcomes unless dispositional entitlement was present.

In Study 3, the AEPQ was positively correlated with both subscales of Chowning and Campbell's (2009) measure of academic entitlement and, as hypothesized, the AEPQ was negatively related to grade fairness. However, only the AEPQ predicted a statistically significant amount of unique variance in grade fairness in a multiple regression also containing Chowning and Campbell's two-component measure of academic entitlement. Both Chowning and Campbell's measure and the AEPQ occupy the same construct space, but when included together the overlap is such that only the AEPQ explains unique variance. With this in mind, the AEPQ better predicted grade fairness than Chowning and Campbell's (2009) two components of academic entitlement when both were present in the model.

Strengths of this study and the results include samples of students with a variety of majors, representation of diverse racial groups (e.g., the large Hispanic representation), and the use of both response schedules with different verbal anchors and different numeric responses to reduce common method variance or common source bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff, MacKenzie, & Podsakoff, 2012). Because Chowning and Campbell (2009) suggested that academic entitlement may ease over time and focused their study mainly on first-year students, the current study used participants from third-year courses at the university and found evidence that problems associated with domain-specific entitlement may still exist in these somewhat older college students, especially

given that age was uncorrelated with scores on the AEPQ in all three samples. Although the AEPQ predicted only a small amount of unique variance in university satisfaction or grade fairness, there are many other possible sources of variance. For example, students may be dissatisfied with their university because of the cost of tuition, the distance of their commute, their relationships with roommates, on-campus services, or something as seemingly mundane as the appearance of the campus and its buildings. Their perceptions of grade fairness can be affected by the distribution of the grades in their courses, whether a grading rubric and clear instructions on assignments is provided, the length of time for grades to be posted by the professor, etc. Because the AEPQ measures a narrowly defined disposition that is not likely to be malleable or manipulated (unlike the above examples of other predictors of university satisfaction and grade fairness), yet it still explained unique variance in these two criteria, this suggests that the AEPQ is an appropriate measure of an important construct.

Like many self-report measures, the AEPQ may suffer from some socially desirable responding; in addition, in this study there may have been common method variance. However, data were collected anonymously, thereby helping offset some contamination of the data by social desirability (Paulhus, 1986, 1988). Although common method variance can result in artificially inflated bivariate correlations, none of the variables measured in these three samples were overly collinear. That is, all correlations were less than $|.85|$ (Kline, 2005).

Future researchers should examine the convergent and discriminant validity of the AEPQ and, because of the apparent interchangeability of entitlement and narcissism by researchers (Twenge, 2006; Twenge, *et al.*, 2010), all of the subscales of the Narcissistic Personality Inventory (Raskin & Hall, 1979). Because the Narcissistic Personality Inventory measures several constructs (e.g., entitlement, superiority, vanity) it is likely to show differential validity in the prediction of key outcomes. That is, subscales of the Narcissistic Personality Inventory have distinct correlations with external variables, some of which are adaptive and some that are maladaptive (Emmons, 1987). Because of the multidimensionality of the Narcissistic Personality Inventory, changes in subscale scores can offset each other, making changes to overall Narcissistic Personality Inventory scores difficult to interpret (Trzesniewski, *et al.*, 2008). Research should proceed via the use of the various subscales of the Narcissistic Personality Inventory instead of overall scores to assess where true differences between entitlement and narcissism exist. An examination of the incremental validity of the AEPQ items when analyzed with other measures of academic entitlement (e.g., Achacoso, 2002; Greenberger, *et al.*, 2008; Kopp, *et al.*, 2011) in the prediction of outcomes like those in the current study might

also prove fruitful (Benson, 1998). The AEPQ may predict dysfunctional behaviors like academic dishonesty and classroom behavior problems, because these behaviors tend to be a reaction to the perceived unfairness (Adams, 1963, 1965) which entitled persons experience (Huseman, *et al.*, 1985, 1987). With these possibilities in mind, it is clear that much is yet to be discovered regarding the measurement of academic entitlement and its place in the nomological network, but that the use of an equity theory-based framework has merit.

REFERENCES

- ACHACOSO, M. V. (2002) What do you mean my grade is not an A? An investigation of academic entitlement, causal attributions, and self-regulation in college students. Unpublished doctoral dissertation, The University of Texas, Austin. Retrieved from <http://repositories.lib.utexas.edu/handle/2152/2486>.
- ADAMS, J. S. (1963) Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67, 422-436.
- ADAMS, J. S. (1965) Inequity in social exchange. In R. M. Steers & L. W. Porter (Eds.), *Motivation and work behavior*. New York: McGraw-Hill, Inc. Pp. 138-154.
- BAUMEISTER, R. F., CAMPBELL, J. D., KRUEGER, J. I., & VOHS, K. D. (2005) Exploding the self-esteem myth. *Scientific American*, 292(1), 84-91.
- BENSON, J. (1998) Developing a strong program of construct validation: a test anxiety example. *Educational Measurement Issues and Practice*, 17, 10-17.
- BENTLER, P. M. (1990) Comparative fit indexes in structural equation models. *Psychological Bulletin*, 107, 238-246.
- BENTLER, P. M. (1998, 10 March) Kurtosis, residuals, fit indices [Msg 011264]. Message posted to SEMNET listserv, archived at <http://bama.ua.edu/cgi-bin/wa?A2=ind9803&L=semnet&P=R10144&l=1>.
- BENTLER, P. M., & WU, E. J. (2002) EQS for Windows user's guide. Encino, CA: Multivariate Software.
- BERGMAN, J. Z., WESTERMAN, J. W., & DALY, J. P. (2010) Narcissism in Management Education. *Academy of Management Learning & Education*, 9, 119-131.
- BOGART, L. M., BENOTSCH, E. G., & PAVLOVIC, J. D. (2004) Feeling superior but threatened: the relation of narcissism to social comparison. *Basic and Applied Social Psychology*, 26, 35-44.
- BRANDEN, N. (1987) How to raise your self-esteem: the proven action-oriented approach to greater self-respect and self-confidence. New York: Bantam Books.
- BRANDEN, N. (1994) The six pillars of self-esteem. New York: Bantam Books.
- BROWNE, M. W., & CUDECK, R. (1993) Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage. Pp. 445-455.
- CAMPBELL, W. K., BONACCI, A. M., SHELTON, J., EXLINE, J. J., & BUSHMAN, B. J. (2004) Psychological entitlement: interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment*, 83, 29-45.
- CHOWNING, K., & CAMPBELL, N. J. (2009) Development and validation of a measure of academic entitlement: individual differences in students' externalized responsibility and entitled expectations. *Journal of Educational Psychology*, 101, 982-997.

- COHEN-CHARASH, Y., & SPECTOR, P. E. (2001) The role of justice in organizations: a meta-analysis. *Organizational Behavior and Human Decision Processes*, 86, 278-321.
- COLQUITT, J. A. (2001) On the dimensionality of organizational justice: a construct validation of a measure. *Journal of Applied Psychology*, 86, 386-400.
- COLQUITT, J. A., CONLON, D. E., WESSON, M. J., PORTER, C. O., & NG, K. Y. (2001) Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86, 425-445.
- COVINGTON, M. V., & MUELLER, K. J. (2001) Intrinsic versus extrinsic motivation: an approach/avoidance reformulation. *Educational Psychology Review*, 13(2), 157-176.
- DECARLO, L. T. (1997) On the meaning and use of kurtosis. *Psychological Methods*, 2, 292-307.
- EMMONS, R. A. (1987) Narcissism: theory and measurement. *Journal of Personality and Social Psychology*, 52, 11-17.
- EPSTEIN, S. (1979) Stability of behavior: on predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097-1126.
- FESTINGER, L. (1954) A theory of social comparison processes. *Human Relations*, 7, 117-140.
- FESTINGER, L. (1957) *A theory of cognitive dissonance*. Stanford, CA: Stanford Univer. Press.
- FINE, R. (1986) *Narcissism, the self and society*. New York: Columbia Univer. Press.
- FLEESON, W. (2004) Moving personality beyond the person-situation debate: the challenge and opportunity of within-person variability. *Current Directions in Psychological Science*, 13, 83-87.
- FOOTE, D. A., & HARMON, S. (2006) Measuring equity sensitivity. *Journal of Managerial Psychology*, 21, 90-108.
- FREUD, S. (1961) The ego and the id. In J. Strachey (Ed. and Trans.), *The standard edition of the complete works of Sigmund Freud*. Vol. 19. London, UK: Hogarth Press. Pp. 12-66. [Original work published 1923]
- GREENBERGER, E., LESSARD, J., CHEN, C., & FARRUGGIA, S. P. (2008) Self-entitled college students: contributions of personality, parenting, and motivation factors. *Journal of Youth and Adolescence*, 37, 1193-1204.
- HACKMAN, J. R., & OLDFHAM, G. R. (1975) Development of the job diagnostic survey. *Journal of Applied Psychology*, 60, 159-170.
- HENSON, R. K. (1999) Multivariate normality: what is it and how is it assessed? In B. Thompson (Ed.), *Advances in social science methodology*. Vol. 5. Stamford, CT: JAI Press. Pp. 193-211.
- HU, L., & BENTLER, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- HUSEMAN, R. C., HATFIELD, J. D., & MILES, E. W. (1985) Test for individual perceptions of job equity: some preliminary findings. *Perceptual & Motor Skills*, 61, 1055-1064.
- HUSEMAN, R. C., HATFIELD, J. D., & MILES, E. W. (1987) A new perspective on equity theory: the equity sensitivity construct. *Academy of Management Review*, 12, 222-234.
- JÖRESKOG, K., & SÖRBOM, D. (2006) *LISREL 8.80*. Chicago, IL: Scientific Software International. [Computer software]
- JUDGE, T. A., EREZ, A., BONO, J. E., & THORESEN, C. J. (2003) The Core Self-Evaluations Scale: development of a measure. *Personnel Psychology*, 56, 303-331.

- KING, W. C., JR., MILES, E. W., & DAY, D. D. (1993) A test and refinement of the equity sensitivity construct. *Journal of Organizational Behavior*, 14, 301-317.
- KLINE, R. B. (2005) *Principles and practices of structural equation modeling*. (2nd ed.) New York: Guilford.
- KOPP, J. P., ZINN, T. E., FINNEY, S. J., & JURICH, D. P. (2011) The development and evaluation of the academic entitlement questionnaire. *Measurement and Evaluation in Counseling and Development*, 44, 105-129.
- LASCH, C. (1978) *The culture of narcissism: American life in an age of diminishing expectations*. New York: W. W. Norton.
- MECCA, A. M., SMELSER, N. J., & VASCONCELLOS, J. (Eds.) (1989) *The social importance of self-esteem*. Berkeley, CA: Univer. of California Press.
- MILLER, B. K. (2009) Confirmatory factor analysis of the equity preference questionnaire. *Journal of Managerial Psychology*, 24, 328-347.
- MILLER, B. K., & NICOLS, K. M. (2011) A domain-specific validation of the core self-evaluations scale. *Psychological Reports*, 109(2), 1-23.
- MILLON, T. (1981) *Disorders of personality*. New York: Wiley.
- MILLON, T. (1998) DSM Narcissistic Personality Disorder. In E. Ronningstam (Ed.), *Disorders of narcissism: diagnostic, clinical, and empirical implications*. Washington, DC: American Psychiatric Press. Pp. 75-101.
- MILLON, T., & DAVIS, R. D. (2000) *Personality disorders in modern life*. New York: John Wiley.
- MOORE, B. E. (1975) Toward a clarification of the concept of narcissism. *Psychoanalytic Study of the Child*, 30, 243-276.
- MOSAK, H. H. (1959) The getting type: a parsimonious social interpretation of the oral character. *Journal of Individual Psychology*, 15, 193-198.
- MOSAK, H. H. (1971) Lifestyle. In A. G. Mikelly (Ed.), *Techniques for behavior change*. Springfield, IL: Thomas. Pp. 77-81.
- NUNNALLY, J. (1978) *Psychometric theory*. (2nd ed.) New York: McGraw-Hill.
- PAULHUS, D. L. (1986) Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: current issues in theory and measurement*. New York: Springer-Verlag. Pp. 143-165.
- PAULHUS, D. L. (1988) Assessing self-deception and impression management in self-reports: the Balanced Inventory of Desirable Responding. Unpublished manual, Univer. of British Columbia, Vancouver, Canada.
- PAULHUS, D. L., & WILLIAMS, K. M. (2002) The dark triad of personality: narcissism, Machiavellianism and psychopathy. *Journal of Research in Personality*, 36, 556-563.
- PODSAKOFF, P. M., MACKENZIE, S. B., LEE, J., & PODSAKOFF, P. N. (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- PODSAKOFF, P. M., MACKENZIE, S. B., & PODSAKOFF, N. P. (2012) Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 65, 539-569.
- PRYOR, L. R., MILLER, J. D., & GAUGHAN, E. T. (2008) A comparison of the psychological entitlement scale and the narcissistic personality inventory's entitlement scale: Relations with general personality traits and personality disorders. *Journal of Personality Assessment*, 90, 517-520.

- RASKIN, R. N., & HALL, C. S. (1979) A narcissistic personality inventory. *Psychological Reports*, 45, 590.
- RASKIN, R. N., & TERRY, H. (1988) A principal-components analysis of the Narcissistic Personality Inventory. *Journal of Personality and Social Psychology*, 54, 890-902.
- RHODEWALT, F., MADRIAN, J. C., & CHENEY, S. (1998) Narcissism, self-knowledge organization, and emotional reactivity: the effect of daily experiences on self-esteem and affect. *Personality and Social Psychology Bulletin*, 24, 75-87.
- RYAN, R. M., & BROWN, K. W. (2006) What is optimal self-esteem? The cultivation and consequences of contingent vs. true self-esteem as viewed from the self-determination theory perspective. In M. H. Kernis (Ed.), *Self-esteem issues and answers: a sourcebook of current perspectives*. New York: Psychology Press. Pp. 125-131.
- SANDLER, J., HOLDER, A., & DARE, C. (1976) Frames of reference in psychoanalytic psychology. *British Journal of Medical Psychology*, 49, 267-274.
- SAULEY, K. S., & BEDEIAN, A. G. (2000) Equity sensitivity: construction of a measure and examination of its psychometric properties. *Journal of Management*, 26, 885-910.
- STOUT, M. (2000) *The feel-good curriculum: the dumbing down of America's kids in the name of self-esteem*. Cambridge, MA: Perseus Publishing.
- THOMPSON, B. (1995) Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement*, 55(4), 525-534.
- TIECHOLZ, J. G. (1978) A selective review of the psychoanalytic literature on theoretical conceptualizations of narcissism. *Journal of the American Psychoanalytic Association*, 26, 831-861.
- TRZESNEWSKI, K. H., DONNELLAN, M. B., & ROBINS, R. W. (2008) Do today's young people really think they are so extraordinary? An examination of secular trends in narcissism and self-enhancement. *Psychological Science*, 19, 181-188.
- TWENGE, J. M. (2006) *Generation me: why today's young Americans are more confident, assertive, entitled – and more miserable than ever before*. New York: Free Press.
- TWENGE, J. M., CAMPBELL, S. M., HOFFMAN, B. J., & LANCE, C. E. (2010) Generational differences in work values: leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of Management*, 36, 1117-1142.
- TWENGE, J. M., KONRATH, S., FOSTER, J. D., CAMPBELL, W. K., & BUSHMAN, B. J. (2008a) Egos inflating over time: a cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, 76, 875-901.
- TWENGE, J. M., KONRATH, S., FOSTER, J. D., CAMPBELL, W. K., & BUSHMAN, B. J. (2008b) Further evidence of an increase in narcissism among college students. *Journal of Personality*, 76, 919-927.
- WEST, S. G., FINCH, J. F., & CURRAN, P. J. (1995) Structural equation models with non-normal variables: problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications*. Thousand Oaks, CA: Sage. Pp. 57-75.
- ZEIGLER-HILL, V., CHADHA, S., & OSTERMAN, L. (2008) Psychological defense and self-esteem instability: is defense style associated with unstable self-esteem? *Journal of Research in Personality*, 42, 348-364.
- ZEIGLER-HILL, V., MYERS, E. M., & CLARK, C. B. (2010) Narcissism and self-esteem reactivity: the role of negative achievement events. *Journal of Research in Personality*, 44, 285-292.

DOOR-IN-THE-FACE: IS IT REALLY NECESSARY THAT BOTH REQUESTS BE MADE BY THE SAME REQUESTER?¹

LOHYD TERRIER

Ecole Hôtelière de Lausanne, Switzerland

BÉNÉDICTE MARFAING

MARC-OLIVIER BOLDI

*Laboratoire de Psychologie
Université de Fribourg, Switzerland*

Ecole Hôtelière de Lausanne, Switzerland

Summary.—The door-in-the-face technique increases the likelihood of individuals accepting a target request by confronting them, beforehand, with an extreme request. The present research tests a new door-in-the-face technique in which the two requests are formulated by two different requesters during the same interaction. 160 participants were asked to help a charity organization following a door-in-the-face procedure. According to the experimental conditions, requests were formulated by one or two requesters during the same or a different interaction. As predicted, the door-in-the-face effect was observed even if two requests are formulated by two requesters, but only if both are present during the interaction. Results are discussed in terms of a reciprocal concessions interpretation and a motivational interpretation.

Introduced by Cialdini, Vincent, Lewis, Catalan, Wheeler, and Darby (1975), the door-in-the-face procedure increases the likelihood that individuals will comply with a target request after turning down a larger request. In Cialdini, *et al.*'s study (1975), researchers wanted college students to accept taking a group of young delinquents to the zoo for two hours. When asked directly, the request was accepted by 16.7% of the students. However, when that request was preceded by the refusal of a highly demanding request (too costly to be accepted), such as working as a volunteer in a juvenile detention center for two days a week during a two-year period, then 50% of the students agreed to the two-hour trip to the zoo (Cialdini, *et al.*, 1975, Exper. 1).

Based on reciprocity norm (Gouldner, 1960) and bargaining conducts (Komorita & Brenner, 1968; Benton, Kelley & Liebling, 1972) studies, Cialdini, *et al.* (1975; Cialdini & Goldstein, 2004) interpreted these results in terms of reciprocal concessions. Shifting from a highly demanding request to a less costly one could be considered as a concession on the requester's part; in response to this concession, the participant would accept the target request.

¹Address correspondence concerning this paper to Lohyd Terrier, Ecole Hôtelière de Lausanne, Le Chalet-à-Gobet, CH-1000 Lausanne 25, Switzerland or e-mail (lohyd.terrier@ehl.ch).

This interpretation is supported by subsequent research (Cialdini, *et al.*, 1975), indicating that the frequency of compliance decreased if the initial request was not demanding enough (Exper. 3), and if the participant did not refuse it (Exper. 1). Additional support comes from findings indicating that the frequency of compliance also decreased when the requests were not made by the same requester (Exper. 2).

Although this explanation has been highlighted by various research (Reeves, Baker, Boyd, & Cialdini, 1991; Cialdini & Goldstein, 2004; Cialdini, 2008; Fennis, Janssen, & Vohs, 2008; Fennis & Janssen, 2010; Fennis & Stel, 2011), it does not seem to be compatible with general results (see O'Keefe & Hale, 1998; Pascual & Guéguen, 2005; Feeley, Anker, & Aloe, 2012). Firstly, it does not enable understanding of why pro-social requests are more efficient than those that are not (Dillard & Hale, 1992; O'Keefe, 1999; Tusing & Dillard, 2000; Millar, 2002). Secondly, according to Abrahams and Bell (1994) and Bell, Abrahams, Clark, and Schlatter (1996), a reciprocal-concession interpretation suggests that the size of the concession should affect the door-in-the-face effects (i.e., the bigger the concession, the bigger the effect). However, Fern, Monroe, and Avila (1986) show that an extreme concession does not necessarily produce more compliance than a moderate one.

In answer to these criticisms, Cialdini and Goldstein (2004) suggested a threshold effect. The latter being the reason for not observing any effects due to the size of the concession. They state that the participant's choice is dichotomous within the door-in-the-face framework: they either accept the choice or reject it. As the gap between both requests reaches a sufficient threshold, it activates the reciprocal concession mechanism, so the only way for the participant to conform would be by accepting the target request. Thus, as soon as the threshold has been reached, the reciprocal-concession mechanism leads the participant to accept the target request, no matter the size of the concession. The type of request used, pro-social versus non pro-social, is also covered by this conceptualization of reciprocal concession. Hale and Laliker (1999) pointed out that the pro-social requests are generally consistent with participants' attitudes. Therefore, one might expect participants to find it easier to accept a request they will have a positive attitude about.

There are additional questions that can be asked regarding why the procedure is not efficient unless both requests are formulated by the same requester (Cialdini, *et al.*, 1975, Exper. 2). As proposed, this postulate supports the reciprocal concession model; however, it should not be relevant to the motivational interpretation suggested by Terrier and Joule (2008) and Terrier, Joule, and Marfaing (2011).

This interpretation is based on studies conducted by Pendleton and Batson (1979) and Tusing and Dillard (2000), respectively emphasizing links between self-presentation, social responsibility norms, and door-in-the-face

effects. It suggests that in this type of situation, the extreme request will activate a social responsibility norm (Gouldner, 1960) and remind the participant of socially valued behavior. Yet, the importance of this request is such that the participant can only refuse it, thereby transgressing an important social norm (Bell, *et al.*, 1996; Tusing & Dillard, 2000). The participant could then fear a negative evaluation, considering that in this particular situation, his refusal would not be acceptable in the eyes of the requester. According to this interpretation, the non-acceptance of this initial refusal could induce a state of tension for the participant that may, as for any state of tension, generate an appeal to its reduction (Terrier & Joule, 2008, Exper. 1). If at the same moment, the requester presents an opportunity to reduce it, in this case by accepting the second less costly request, the participant could take it in order to maintain a positive image of himself.

On this basis, Terrier and Joule (2008, Exper. 2) found that making the first refusal acceptable reduced the state of tension experienced by participants, and consequently the door-in-the-face effect was observed. They also show that the door-in-the-face effect could be increased when the non-compliance with the extreme request is perceived as unacceptable by the requester (Terrier, *et al.*, 2011). These results support the idea that the compliance with the target request is not due to the need to comply with a concession made by the requester, but to the need for participants to restore self-image. Therefore, it can be assumed that the similarity between the requesters is not a necessary condition to obtain door-in-the-face effects.

The main objective of this study is to test a new modality of the door-in-the-face effect, for which the requests are formulated by two different requesters, but in a situation where all the actors are present during both formulations. Here, the reciprocal concession interpretation predicts there should be no door-in-the-face effect, because the participants cannot respond to the first requester's concession. On the other hand, the motivational interpretation predicts a significant door-in-the-face effect, because accepting a second request will allow the participant to redeem himself to the first requester. For classic comparisons, both interpretations predict the same results.

METHOD

Participants

Passers-by ($N=160$; 73 men, 87 women), between 18 and 50 years of age, were approached in a street of Nîmes, France, by two female confederates (all confederates were women, ages 20 to 22 years).

Procedure

This experiment includes three experimental conditions and one control condition. To make comparisons with previous studies easier, the same experimental procedure as Cialdini, *et al.* (1975, Exper. 2) was used. Partici-

pants were approached in a pedestrian street by two confederates (A and B) introducing themselves as members of an animal protection association. At the same time, another confederate (C) arrived and engaged in a conversation with confederate B, whom he appears to know. For the three experimental conditions, A formulated the first request, asking the participants to get actively involved in the association for two years (the extreme request).

Once this extreme request was refused, the participants were randomly assigned to one of the three experimental conditions: (1) Classic Door-in-the-face: in the classic door-in-the-face condition, after the refusal of the extreme request, confederate A formulated the second request, by asking participants to get involved and help the association for one afternoon only (target request; see Appendix). (2) Two requesters–Absent: in this condition, after the refusal of the extreme request, confederate A thanked the participant and left with confederate B, leaving the participant alone with confederate C. The latter, mentioned that he overheard his refusal of the extreme request, then asked participants to get involved by helping the association for one afternoon only (target request). (3) Two requesters–Present: in this condition, confederate A thanked the participant but before his departure, confederate C asked participant to get involved by helping the association for one afternoon only (target request). (4) In the Control condition, confederate A asked the participant to get involved by helping the association for one afternoon only (target request).

Hypothesis 1. A higher rate of acceptance of the target request should occur in the classic door-in-the-face condition than in the control condition.

Hypothesis 2. A lower rate of acceptance of the target request should occur when the two requests are made by two different requesters (Two requesters–Absent) than when they are formulated by the same requester (Classic Door-in-the-face).

Hypothesis 3. A lower rate of acceptance of the target request should occur when the two requests are made by two different requesters in two different interactions (Two requesters–Absent) than when they are formulated by two different requesters in the same interaction (Two requesters–Present).

Hypothesis 4. A higher rate of acceptance of the target request should occur in the condition Two requesters–Present than in the Control condition.

Statistical analyses were conducted in R (Team, 2010). The model is a logistic regression. The dependent variable is the target request (No = 0, Yes = 1) and the independent variable is the experimental conditions factor (Control, Classic Door-in-the-face, Two requesters–Absent, Two requesters–Present).

RESULTS

As in almost all research on the door-in-the-face procedure, all participants refused the extreme request. The classic door-in-the-face effect was observed, supporting Hypothesis 1: the acceptance rate of the target request was higher in the Classic Door-in-the-face condition than in the Control condition: $\text{Wald}(1, N=80)=15.18, p<.001, \phi=0.43$. In accordance with Cialdini, *et al.* (1975) and Hypothesis 2, the door-in-the-face procedure generated significantly more compliance with the target request in the Classic Door-in-the-face condition than in the Two requesters–Absent condition: $\text{Wald}(1, N=80)=15.18, p<.001, \phi=0.43$. Lastly, Hypotheses 3 and 4 were also supported. The target request's acceptance rate was higher for the Two requesters–Present condition than for the Two requesters–Absent condition and for the Control condition: $\text{Wald}(1, N=80)=12.78, p<.001, \phi=0.40$. These results highlight the door-in-the-face effect, whereas both requests are formulated by two requesters, therefore supporting the motivational interpretation (Table 1).

TABLE 1
ACCEPTANCE AND REJECTION RATES FOR EACH CONDITION

Classic Door-in-the-face		Two Requesters–Absent		Two Requesters–Present		Control	
%	Accept/Reject	%	Accept/Reject	%	Accept/Reject	%	Accept/Reject
55	22/40	10	4/40	50	20/40	10	4/40

Note.—No significant difference between the Classic Door-in-the-face and Two requesters–Present.

DISCUSSION

Interpretations of the door-in-the-face effect, in terms of reciprocal concessions (Cialdini, *et al.*, 1975; Cialdini & Goldstein, 2004), depend on a series of experiments showing that this technique produced compliance only if both requests were formulated by the same requester. The current study tested a new experimental situation where predictions based on the standard interpretation of reciprocal concessions opposed those based on a motivational interpretation (Terrier & Joule, 2008; Terrier, *et al.*, 2011). In the situation where both requests are formulated by two different requesters, but in the presence of the first requester, the reciprocal-concessions interpretation and motivational interpretation suggest opposite findings. The first interpretation predicts the absence of the door-in-the-face effect (accepting the target request does not allow the participant to answer the concession made by the requester) and the second predicts the presence of a signif-

icant door-in-the-face effect, because accepting the target request could reduce the discomfort generated by the refusal of the extreme request. The results of the present experiment clearly support the motivational hypothesis, highlighting the door-in-the-face effect whereby two requests are formulated by two different requesters.

According to the motivational interpretation, refusing the first request is perceived by the participant as unacceptable for the requester, generating a state of tension induced by self-image (Terrier & Joule, 2008; Terrier, *et al.*, 2011). Therefore, accepting the target request enables the participant to reduce this tension by restoring her image in the eyes of the requester. Here, it is not the acceptance of the second target that is considered important, but reducing the tension created by the initial refusal. Accepting the second request while the first requester is still present allows the participant to reduce that tension. In contrast, interpretations based on reciprocal concession do not lead to these results. So in this case, although the first requester is present, the participant does not have the opportunity to return the concession. Therefore, just like in the Two requesters–Absent condition, the door-in-the-face effect should not occur.

This study also supports results reached by O'Keefe and Figge (1997) and Feeley, *et al.* (2012) nuancing the question related to the requester. According to them, one should not confuse the requester with the beneficiary: in Cialdini, *et al.*'s (1975) experiment, for example, the requester is the confederate, and the beneficiary is a detention center. O'Keefe and Figge (1997) made some complementary analyses of Dillard, Hunter, and Burgoon's (1984) meta-analysis to compare the effects of the technique, more precisely when both requests are beneficial to one or two different parties. It appears the door-in-the-face effect is greater if the same beneficiary is benefitted by both requests. In the current case, the beneficiary is the same, but the requester changes. It would be interesting to test this new procedure in a situation where the beneficiaries of the two requests are different.

Finally, if these results are in contradiction with a reciprocal concessions interpretation, they do support the hypothesis linked to social responsibility (Bell, *et al.*, 1996; Tusing & Dillard, 2000), self-presentation (Pendleton & Batson, 1979), and guilt (O'Keefe & Figge, 1997). Each of these studies postulates a psychological state that may be accompanied by a state of tension. The theoretical interpretation in this study allows integration of all three interpretations within the same motivational framework. Further research is needed to distinguish the relevance of each of these, including manipulating the similarity of the beneficiaries (here, motivational interpretation and social responsibility interpretation could provide two opposite hypotheses).

In conclusion, if this study clearly states that the motivational interpretation allows better understanding of the results obtained in the present research, further research should test how it could help recognize all the results found using the door-in-the-face technique.

REFERENCES

- ABRAHAMS, M., & BELL, R. (1994) Encouraging charitable contributions: an examination of three models of door-in-the-face compliance. *Communication Research*, 21, 131-153.
- BELL, R. A., ABRAHAMS, M. F., CLARK, C. L., & SCHLATTER, C. (1996) The door-in-the-face compliance strategy: an individual differences analysis of two models in an AIDS fundraising context. *Communication Quarterly*, 44(1), 107-124.
- BENTON, A. A., KELLEY, H. H., & LIEBLING, B. (1972) Effects of extremity of offers and concession rate on the outcomes of bargaining. *Journal of Personality and Social Psychology*, 24(1), 73-83.
- CIALDINI, R. B. (2008) *Influence: science and practice*. (5th ed.) New-York: Quill, Prentice Hall.
- CIALDINI, R. B., & GOLDSTEIN, N. J. (2004) Social influence: compliance and conformity. *Annual Review of Psychology*, 55, 591-621.
- CIALDINI, R. B., VINCENT, J. E., LEWIS, S. K., CATALAN, J., WHEELER, D., & DARBY, B. L. (1975) A reciprocal concessions procedure for inducing compliance: the door-in-the-face technique. *Journal of Personality and Social Psychology*, 31, 206-215.
- DILLARD, J., & HALE, J. (1992) Prosocialness and sequential request compliance techniques: limits to the foot-in-the-door and the door-in-the-face. *Communication Studies*, 43, 220-232.
- DILLARD, J. P., HUNTER, J. E., & BURGOON, M. (1984) Sequential-request persuasive strategies: meta-analysis of foot-in-the-door and door-in-the-face. *Human Communication Research*, 10, 461-488.
- FEELEY, T. H., ANKER, A. E., & ALOE, A. M. (2012) The door-in-the-face persuasive message strategy: a meta-analysis of the first 35 years. *Communication Monographs*, 79(3), 316-343.
- FENNIS, B., & JANSSEN, L. (2010) Mindlessness revisited: sequential request techniques foster compliance by draining self-control resources. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 29(3), 235-246.
- FENNIS, B. M., JANSSEN, L., & VOHS, K. D. (2008) Acts of benevolence: a limited resources account of compliance with charitable requests. *Journal of Consumer Research*, 35, 906-924.
- FENNIS, B. M., & STEL, M. (2011) The pantomime of persuasion: fit between nonverbal communication and influence strategies. *Journal of Experimental Social Psychology*, 47, 806-810.
- FERN, E., MONROE, K., & AVILA, R. (1986) Effectiveness of multiple request strategies: a synthesis of research results. *Journal of Marketing Research*, 24, 64-80.
- GOULDNER, A. W. (1960) The norm of reciprocity: a preliminary statement. *American Sociological Review*, 25, 161-178.
- HALE, J. J., & LALIKER, M. (1999) Explaining the door-in-the-face: is it really time to abandon reciprocal concessions? *Communication Studies*, 50(3), 203-210.
- KOMORITA, S. S., & BRENNER, A. R. (1968) Bargaining and concession making under bilateral monopoly. *Journal of Personality and Social Psychology*, 9(1), 15-20.

- MILLAR, M. G. (2002) The effectiveness of the door-in-the-face compliance strategy on friends and strangers. *Journal of Social Psychology*, 142(3), 295-304.
- O'KEEFE, D. J. (1999) Three reasons for doubting the adequacy of the reciprocal concessions explanation of door-in-the-face effects. *Communication Studies*, 50(3), 211-220.
- O'KEEFE, D. J., & FIGGE, M. (1997) A guilt-based explanation of the door-in-the-face influence strategy. *Human Communication Research*, 24, 64-81.
- O'KEEFE, D. J., & HALE, S. L. (1998) The door-in-the-face influence strategy: a random-effects meta-analytic review. *Communication Yearbook*, 21, 1-33.
- PASCUAL, A., & GUÉGUEN, N. (2005) Foot-in-the-door and door-in-the-face: a comparative meta-analytic study. *Psychological Reports*, 96(1), 122-128.
- PENDLETON, M. G., & BATSON, C. D. (1979) Self-presentation and the door-in-the-face technique for inducing compliance. *Personality and Social Psychology Bulletin*, 5, 77-91.
- REEVES, R. A., BAKER, G. A., BOYD, J. G., & CIALDINI, R. B. (1991) The door-in-the-face technique: reciprocal concessions vs. self-presentational explanations. *Journal of Social Behavior & Personality*, 6, 545-558.
- TEAM, R. D. C. (2010) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- TERRIER, L., & JOULE, R. V. (2008) La procédure de porte-au-nez: vers une interprétation motivationnelle [The door-in-the-face technique: toward a motivational interpretation]. *Cahiers Internationaux de Psychologie Sociale*, 77, 5-14.
- TERRIER, L., JOULE, R. V., & MARFAING, B. (2011) Requester's acceptance and non-acceptance of the refusal of the initial request: how to improve the door-in-the-face effects? *Current Research in Social Psychology*, 17(1), 1-9.
- TUSING, K. J., & DILLARD, J. P. (2000) The psychological reality of the door-in-the-face: it's helping, not bargaining. *Journal of Language and Social Psychology*, 19, 5-25.

Accepted October 11, 2013.

APPENDIX

DOOR-IN-THE-FACE REQUESTS

Extreme request: I am a member of the "happiness with four legs" association. We welcome cats and dogs that have been abandoned by their masters. We are currently looking for volunteers to help the association on a daily basis. Members' activities mainly consist of cleaning the boxes, feeding the animals, walking the dogs, and taking care of them. We need reinforcements who would agree to be present on Saturday afternoons to walk the dogs and be ready to commit for at least two years. Would you be interested?

Target request: We are also looking for occasional help, such as volunteers that would agree to take the dogs on a walk, one Saturday afternoon within the next following weeks. Would you be interested?

ERRATUM

VIVES, J., LOSILLA, J-M., RODRIGO, M-F, PORTELL, M., & LLORENS, N. (2008) Overdispersion tests in count-data analysis. *Psychological Reports*, 103, 145-160. DOI: 10.2466/PR0.103.1.145-160

The originally published authors of this article report that, by court order (Ruling No. 235/12), Noelia Llorens must be added as a fifth author. The above corrected citation reflects this change.

SECTION 1: DISABILITY & TRAUMA

<i>Target-to-distractor Ratio Effects on Decision Time in the Orderly Array Shape Cancellation Task:</i> YOSHIMI NAKAJIMA, YOSHIFUMI IKEDA, AND HIDEYUKI OKUZUMI	353
--	-----

SECTION 2: EMPLOYMENT PSYCHOLOGY & MARKETING

<i>Validation of the Karasek–Job Content Questionnaire to Measure Job Strain in Vietnam:</i> THI GIANG HOANG, MARC CORBIÈRE, ALESSIA NEGRINI, MINH KHUÊ PHAM, AND DANIEL REINHARZ	363
<i>Psychometric Properties of the Achievement Goal Questionnaire–Revised Administered to Greek University Students:</i> MARY APOSTOLOU	380
<i>Brand Emotional Credibility: Effects of Mixed Emotions About Branded Products with Varying Credibility:</i> ANTONIO MILETI, M. IRENE PRETE, AND GIANLUIGI GUIDO	404
<i>Predicting Game-attending Behavior in Amateur Athletes: The Moderating Role of Intention Stability:</i> WAN CHEN LU, CHIH-FU CHENG, AND LUNG HUNG CHEN	420

SECTION 3: MEASURES & STATISTICS

<i>Levels Of Recovery Scale (LORS): Psychometric Properties of a New Instrument to Assess Psychotic Symptoms and Patient Awareness:</i> DONALD P. CORRIVEAU AND SHARON SOUSA	435
<i>Factorial Structure and Convergent and Discriminant Validity of the E (Empathy) Scale:</i> ULRICH S. TRAN, ANTON-RUPERT LAIREITER, DAVID P. SCHMITT, CHRISTINE NEUNER, MAX LEIBETSEDER, SARA LEYLA SZENTE-VORACEK, AND MARTIN VORACEK	441
<i>Generalizability of Delay of Gratification: Dimensionality and Function:</i> XIAOYAN LIU, LEI WANG, AND JING JIANG	464
<i>Gender Invariance of the College Student Stress Scale:</i> RONALD C. FELDT AND CHRISTINA UPDEGRAFF	486
<i>Teacher Stress Questionnaire: Validity and Reliability Study in Italy:</i> MARIA CLELIA ZURLO, DANIELA PES, AND ROBERTO CAPASSO	490

SECTION 4: MENTAL & PHYSICAL HEALTH

<i>National Intelligence Estimates and the Failed State Index:</i> MARTIN VORACEK	519
<i>Perceived Likelihood of Experiencing a Desirable Versus Undesirable Outcome:</i> RON S. GOLD	525
<i>Object Exploration in 3- To 7-year-old Children:</i> MAGDALENA ŚNIEGULSKA AND WOJCIECH PISULA	528
<i>Effects of Prolonged Wakefulness: The Role of PERIOD3 Genotypes and Personality Traits:</i> GIUSEPPE BARBATO, ANTONIO COSTANZO, CIRO DELLA MONICA, PAOLO D'ONOFRIO, FLAVIA CERRATO, AND VITTORIA DE PADOVA	540
<i>Up- and Down-regulation of Daily Emotion: An Experience Sampling Study of Chinese Adolescents' Regulatory Tendency and Effects:</i> XINMEI DENG, BIAO SANG, AND ZIYAN LUAN	552
<i>Longitudinal Predictions of the Brooding and Reflection Subscales of the Japanese Ruminative Responses Scale for Depression:</i> AKIRA HASEGAWA, MUNENAGA KODA, YOSUKE HATTORI, TSUYOSHI KONDO, AND JUN KAWAGUCHI	566
<i>The Structure of Prosocial Behavior: A Comment on Richaud, Mesurado, and Cortada (2013):</i> MARIOLA ŁAGUNA AND PIOTR OLEŚ	586

SECTION 5: RELATIONSHIPS & COMMUNICATIONS

<i>The Working Alliance in Group Counseling: An Exploratory Study:</i> ROSTYSLAW W. ROBAK, KELSEY A. KANGOS, SHEILA H. CHIFFRILLER, AND PAUL W. GRIFFIN	591
<i>Romantic Partners in a Market Perspective: Expectations About What Ensures a Highly Desirable Partner:</i> FELIPE N. CASTRO, WALLISEN T. HATTORI, MARIA EMÍLIA YAMAMOTO, AND FÍVIA A. LOPES	605
<i>Social Support, Work-Family Conflict, and Emotional Exhaustion in South Korea:</i> SOOJIN LEE, SECKYOUNG LORETTA KIM, EUN KYUNG PARK, AND SEOKHWA YUN	619

SECTION 6: SOCIOCULTURAL ISSUES IN PSYCHOLOGY

<i>People's Preference Patterns for Gains/Losses in Multiple Time Period Situations:</i> SHIN-SHIN CHANG AND JUNG-HUA CHANG	635
<i>Why Do Physicians Treat Their Relatives? Exploring the Influence of Social Psychology:</i> JONATHAN R. SCARFF	647
<i>Measurement of Academic Entitlement:</i> BRIAN K. MILLER	654
<i>Door-in-the-face: Is it Really Necessary that Requests Be Made by the Same Requester?</i> LOHYD TERRIER, BÉNÉDICTE MARFAING, AND MARC-OLIVIER BOLDI	675
<i>Erratum</i>	683

